Limiting distribution of the sample canonical correlation coefficients of high-dimensional random vectors

Fan Yang *[1]

[1]Department of Statistics, University of Pennsylvania

October 1, 2021

Abstract

In this paper, we prove a CLT for the sample canonical correlation coefficients between two high-dimensional random vectors with finite rank correlations. More precisely, consider two random vectors $\widetilde{\mathbf{x}} = \mathbf{x} + A\mathbf{z}$ and $\widetilde{\mathbf{y}} = \mathbf{y} + B\mathbf{z}$, where $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{y} \in \mathbb{R}^q$ and $\mathbf{z} \in \mathbb{R}^r$ are independent random vectors with i.i.d. entries of mean zero and variance one, and $A \in \mathbb{R}^{p\times r}$ and $B \in \mathbb{R}^{q\times r}$ are two arbitrary deterministic matrices. Given n samples of $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$, we stack them into two matrices $\mathcal{X} = X + AZ$ and $\mathcal{Y} = Y + BZ$, where $X \in \mathbb{R}^{p\times n}$, $Y \in \mathbb{R}^{q\times n}$ and $Z \in \mathbb{R}^{r\times n}$ are random matrices with i.i.d. entries of mean zero and variance one. Let $\widetilde{\lambda}_1 \geqslant \widetilde{\lambda}_2 \geqslant \cdots \geqslant \widetilde{\lambda}_r$ be the largest r eigenvalues of the sample canonical correlation (SCC) matrix $\mathcal{C}_{\mathcal{X}\mathcal{Y}} = (\mathcal{X}\mathcal{X}^\top)^{-1/2}\mathcal{X}\mathcal{Y}^\top(\mathcal{Y}\mathcal{Y}^\top)^{-1}\mathcal{Y}\mathcal{X}^\top(\mathcal{X}\mathcal{X}^\top)^{-1/2}$, and let $t_1 \geqslant t_2 \geqslant \cdots \geqslant t_r$ be the squares of the population canonical correlation coefficients between $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$. Under certain moment assumptions, we show that there exists a threshold $t_c \in (0,1)$ such that if $t_i > t_c$, then $\sqrt{n}(\widetilde{\lambda}_i - \theta_i)$ converges weakly to a centered normal distribution, where θ_i is a fixed outlier location determined by t_i. Our proof uses a self-adjoint linearization of the SCC matrix and a sharp local law on the inverse of the linearized matrix.

1 Introduction

Given two random vectors $\widetilde{\mathbf{x}} \in \mathbb{R}^p$ and $\widetilde{\mathbf{y}} \in \mathbb{R}^q$, canonical correlation analysis (CCA) has been one of the most classical methods to study the correlations between them since the seminal work by Hotelling [22]. More precisely, CCA seeks two sequences of orthonormal vectors, such that the projections of $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ onto these vectors have maximized correlations. These correlations are referred to as *canonical correlation coefficients* (CCCs), which can be characterized as the square roots of the eigenvalues of the population canonical correlation (PCC) matrix

$$\widetilde{\mathbf{\Sigma}} := \Sigma_{xx}^{-1/2}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1/2},$$

where Σ_{xx}, Σ_{yy}, Σ_{xy} and Σ_{yx} are the population covariance and cross-covariance matrices defined by

$$\Sigma_{xx} := \mathbb{E}(\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}^\top) - (\mathbb{E}\widetilde{\mathbf{x}})(\mathbb{E}\widetilde{\mathbf{x}})^\top, \quad \Sigma_{yy} := \mathbb{E}(\widetilde{\mathbf{y}}\widetilde{\mathbf{y}}^\top) - (\mathbb{E}\widetilde{\mathbf{y}})(\mathbb{E}\widetilde{\mathbf{y}})^\top, \quad \Sigma_{xy} = \Sigma_{yx}^\top := \mathbb{E}(\widetilde{\mathbf{x}}\widetilde{\mathbf{y}}^\top) - (\mathbb{E}\widetilde{\mathbf{x}})(\mathbb{E}\widetilde{\mathbf{y}})^\top.$$

In this paper, we consider the following standard signal-plus-noise model for $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$:

$$\widetilde{\mathbf{x}} = \mathbf{x} + A\mathbf{z}, \quad \widetilde{\mathbf{y}} = \mathbf{y} + B\mathbf{z}, \tag{1.1}$$

*E-mail: fyang75@wharton.upenn.edu.

This work was partially supported by the Wharton Dean's Fund for Postdoctoral Research.

where $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$ are two independent noise vectors with i.i.d. entries of mean zero and variance one, $\mathbf{z} \in \mathbb{R}^r$ is a shared signal vector with i.i.d. entries of mean zero and variance one (which yields a rank-r correlation), and $A \in \mathbb{R}^{p\times r}$ and $B \in \mathbb{R}^{q\times r}$ are two arbitrary deterministic matrices. Under the model (1.1), the PCC matrix is given by a rank-r matrix

$$\widetilde{\mathbf{\Sigma}} = (I_p + AA^\top)^{-1/2}AB^\top(I_p + BB^\top)^{-1}BA^\top(I_p + AA^\top)^{-1/2},$$

and we denote the r non-trivial eigenvalues of $\widetilde{\mathbf{\Sigma}}$ as $t_1 \geqslant t_2 \geqslant \cdots \geqslant t_r \geqslant 0$.

We can study $\widetilde{\mathbf{\Sigma}}$ and the population CCCs via their sample counterparts, i.e., the *sample canonical correlation* (SCC) matrix and the sample CCCs. More precisely, let $(\widetilde{\mathbf{x}}_i, \widetilde{\mathbf{y}}_i)$, $1 \leqslant i \leqslant n$, be n i.i.d. samples of $(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}})$. We stack them (as column vectors) into two matrices

$$\mathcal{X} := n^{-1/2}\left(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \cdots, \widetilde{\mathbf{x}}_n\right) = X + AZ, \quad \mathcal{Y} := n^{-1/2}\left(\widetilde{\mathbf{y}}_1, \widetilde{\mathbf{y}}_2, \cdots, \widetilde{\mathbf{y}}_n\right) = Y + BZ, \tag{1.2}$$

where $n^{-1/2}$ is a convenient scaling, with which we can write the sample covariance and cross-covariance matrices concisely as

$$\widetilde{S}_{xx} := \mathcal{X}\mathcal{X}^\top, \quad \widetilde{S}_{yy} := \mathcal{Y}\mathcal{Y}^\top, \quad \widetilde{S}_{xy} = \widetilde{S}_{yx}^\top := \mathcal{X}\mathcal{Y}^\top,$$

and X, Y and Z are respectively $p \times n$, $q \times n$ and $r \times n$ matrices with i.i.d. entries of mean zero and variance n^{-1}. Then we define the SCC matrix as

$$\mathcal{C}_{\mathcal{X}\mathcal{Y}} := \widetilde{S}_{xx}^{-1/2}\widetilde{S}_{xy}\widetilde{S}_{yy}^{-1}\widetilde{S}_{yx}\widetilde{S}_{xx}^{-1/2}$$

and denote their eigenvalues as $\widetilde{\lambda}_1 \geqslant \widetilde{\lambda}_2 \geqslant \cdots \geqslant \widetilde{\lambda}_{p\wedge q} \geqslant 0$. The square roots of these eigenvalues are referred to as *sample canonical correlation coefficients*. Equivalently, the sample CCCs are the cosines of the principal angles between the two subspaces spanned by the rows of $\mathcal{X}$ and $\mathcal{Y}$, respectively. If $n \to \infty$ while p, q and r are fixed, it is easy to see that the SCC matrix converges to the PCC matrix almost surely by the law of large numbers, and hence every sample CCC converges almost surely to the corresponding population CCC. On the other hand, in this paper, we focus on the high-dimensional setting with a low-rank signal: $p/n \to c_1$ and $q/n \to c_2$ as $n \to \infty$ for some constants $c_1 \in (0, 1)$ and $c_2 \in (0, 1 - c_1)$, and r is a fixed integer that does not depend on n. In this case, the behavior of the SCC matrix deviates greatly from that of the PCC matrix.

Related work. In the null case with $r = 0$, the eigenvalue statistics of the SCC matrix have been well-understood. If $\mathcal{X}$ and $\mathcal{Y}$ are Gaussian matrices, then the eigenvalues of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ reduce to those of a double Wishart matrix, which belongs to the famous Jacobi ensemble [24]. It was shown in [37] that, almost surely, the empirical spectral distribution (ESD) of the double Wishart matrix converges weakly to a deterministic probability distribution (cf. (2.15) below). By analyzing the joint eigenvalue density of the Jacobi ensemble, Johnstone [24] proved that the largest eigenvalues of double Wishart matrices satisfy the Tracy-Widom law asymptotically. Alternatively, the Tracy-Widom law of double Wishart matrices can also be obtained as a consequence of the results in [21] for F-type matrices. In the general non-Gaussian case, the convergence of the ESD of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ was proved in [42], the CLT of the linear spectral statistics for $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ was proved in [43], and the Tracy-Widom law of the largest eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ was proved in [20] under the assumption that the entries of $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ have finite moments up to any order. The moment assumptions for the Tracy-Widom law was later relaxed to the finite fourth moment condition in [40].

Some arguments in the literature for the null case are based on the fact that the subspaces spanned by the rows of $\mathcal{X}$ and $\mathcal{Y}$ are approximately uniform (Haar) distributed random subspaces, which, however, does not hold for the non-null case with $r > 0$. This makes the study of the non-null case more challenging. If $\mathcal{X}$ and $\mathcal{Y}$ are Gaussian matrices, then the asymptotic distributions of the sample CCCs have been derived in the case where either p or q is fixed as $n \to \infty$ [19]. When p and q are both proportional to n, the limiting distributions of the sample CCCs have been established under the Gaussian assumption in [4], which we discuss in more detail now.

BBP transition. Suppose X, Y and Z are independent random matrices with i.i.d. Gaussian entries. Bao et al. [4] proved that for any $1 \leqslant i \leqslant r$, the behavior of $\widetilde{\lambda}_i$ undergoes a sharp transition across the threshold t_c defined by

$$t_c := \sqrt{\frac{c_1 c_2}{(1-c_1)(1-c_2)}}. \tag{1.3}$$

More precisely, the following dichotomy occurs:

(1) if $t_i < t_c$, then $\widetilde{\lambda}_i$ sticks to the right edge λ_+ (cf. (2.16) below) of the limiting bulk eigenvalue spectrum of the SCC matrix, and $n^{2/3}(\widetilde{\lambda}_i - \lambda_+)$ converges weakly to the Tracy-Widom law;

(2) if $t_i > t_c$, then $\widetilde{\lambda}_i$ lies around a fixed location $\theta_i \in (\lambda_+, 1)$, and $n^{1/2}(\widetilde{\lambda}_i - \theta_i)$ converges weakly to a centered normal random variable.

Following the notation in random matrix theory literature, we call $\widetilde{\lambda}_i$ in case (2) an *outlier*. The above abrupt change of the behavior of $\widetilde{\lambda}_i$ when t_i crosses t_c is generally referred to as a *BBP transition*, which dates back to the seminal work of Baik, Ben Arous and Péché [2] on spiked sample covariance matrices. The phenomenon of BBP transition has been observed in many random matrix ensembles deformed by low-rank perturbations. Without attempting to be comprehensive, we refer the reader to [10, 11, 16, 26, 27, 33] about deformed Wigner matrices, [1, 2, 3, 9, 17, 23, 32] about spiked sample covariance matrices, [12, 39, 41] about spiked separable covariance matrices, and [5, 6, 7, 38] about several other types of deformed random matrix ensembles. The SCC matrix $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ considered in this paper can be regarded as a low-rank perturbation of the SCC matrix in the null case with $r = 0$.

Main results and basic ideas. A natural question is whether the above BBP transition holds universally if we only assume certain moment conditions on the entries of X, Y and Z. Answering this question is not only theoretically interesting from the point of view of random matrix theory, but also crucial for modern applications of CCA in e.g., statistical learning, wireless communications, financial economics and population genetics. In this paper, we solve this problem and prove that the BBP transition occurs as long as the entries of X and Y satisfy the bounded $(8+\varepsilon)$-th moment condition (with ε denoting an arbitrarily small positive constant). More precisely, we obtain the following results when $t_i > t_c$.

(i) In Theorem 2.3, assuming that the entries of X, Y and Z have bounded moments up to any order, we prove that $n^{1/2}(\widetilde{\lambda}_i - \theta_i)$ converges weakly to a centered normal random variable.

(ii) In Theorem 2.4, we prove the CLT for $\widetilde{\lambda}_i$ under a relaxed bounded $(8+\varepsilon)$-th moment condition on the entries of X, Y and a bounded $(4+\varepsilon)$-th moment condition on the entries of Z.

On the other hand, when $t_i < t_c$, the Tracy-Widom law of $n^{2/3}(\widetilde{\lambda}_i - \lambda_+)$ was proved in [31]. For the reader's convenience, we will state it in Theorem 2.5.

The proof in [4] depends crucially on the fact that multivariate Gaussian distributions are rotationally invariant under orthogonal transforms, which makes the proof hard to be extended to the non-Gaussian case. To circumvent this issue, we employ an entirely different approach—a linearization method developed in [40]. More precisely, we define a $(p+q+2n) \times (p+q+2n)$ random matrix H that is linear in X and Y (cf. equation (3.6) below) and call its inverse $G := H^{-1}$ as *resolvent*. In [40], we found that the eigenvalues of the SCC matrix $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ are precisely the solutions to a determinant equation in terms of a linear functional of G (cf. equation (3.8) below). Moreover, an (almost) optimal local law for this linear functional was obtained in [40]. In [31], we obtained a large deviation estimate on the outlier sample CCCs—if $t_i > t_c$, then $\widetilde{\lambda}_i$ converges to θ_i with convergence rate $\mathrm{O}(n^{-1/2+\varepsilon})$ (which is slightly larger than the correct order of fluctuation $n^{-1/2}$).

With the local law and the large deviation estimate as main inputs, we can reduce the problem to proving the CLT for a (different) linear functional of G, denoted by $\mathcal{E}(X, Y, Z)$ (cf. Section 3.3). The main technical part of our proof is to show that $\mathcal{E}(X, Y, Z)$ converges weakly to a centered Gaussian random variable. Our basic idea is to use the classical moment method, that is, showing that the moments of $\mathcal{E}(X, Y, Z)$ match those of a Gaussian random variable asymptotically. One method to calculate the moments of $\mathcal{E}(X, Y, Z)$ is to use the simple identity $1 = HG$ and apply a resolvent expansion formula (cf. Lemma B.1 below) to the resulting expression. However, the relevant calculation for this strategy will be rather tedious. Instead, we adopt a strategy in [26, 27], that is, we first prove the CLT in an "almost Gaussian" case (i.e., a case where most of the entries of X and Y are Gaussian), and then show that the general case is sufficiently close to the almost Gaussian case. This strategy allows us to divide the lengthy calculation into several parts that are more manageable. In particular, the resolvent expansion formula can be replaced by a simpler Gaussian integration by parts formula.

Finally, we remark that the limiting variance of $n^{1/2}(\widetilde{\lambda}_i - \theta_i)$ depends on the fourth cumulants of the entries of X, Y and Z in an intricate way, which has not been identified in the Gaussian case. We also perform simulations to verify this deviation from the CLT result in [4] (cf. Figure 1).

Organizations. The rest of this paper is organized as follows. In Section 2, we define the model and state the main results, Theorem 2.3 and Theorem 2.4, on the limiting distributions of the outlier sample CCCs. We will give the proof of Theorem 2.3 in Sections 3–5. In Section 3, we introduce the linearization method, define the resolvent, and reduce the problem to proving the CLT for a linear functional of the resolvent. In Section 4, we establish the CLT of the outlier sample CCCs in an almost Gaussian case, where most of the entries of X and Y are Gaussian. In Section 5, we complete the proof of Theorem 2.3 by showing that the general setting of Theorem 2.3 is close to the almost Gaussian case asymptotically. Finally, utilizing Theorem 2.3 and a comparison argument, we complete the proof of Theorem 2.4 in Section 6. The appendix contains the proofs of Lemma 4.5 and Lemma 4.6, which are two key lemmas in the proof of Theorem 2.3.

Conventions. For two quantities a_n and b_n depending on n, the notation $a_n = \mathrm{O}(b_n)$ means that $|a_n| \leqslant C|b_n|$ for some constant $C > 0$, and $a_n = \mathrm{o}(b_n)$ means that $|a_n| \leqslant c_n|b_n|$ for a positive sequence $c_n \downarrow 0$ as $n \to \infty$. We use the notation $a_n \lesssim b_n$ if $a_n = \mathrm{O}(b_n)$ and the notation $a_n \sim b_n$ if $a_n = \mathrm{O}(b_n)$ and $b_n = \mathrm{O}(a_n)$. Given a matrix A, we use $\|A\| := \|A\|_{l^2 \to l^2}$ to denote the operator norm, $\|A\|_F$ to denote the Frobenius norm, and $\|A\|_{\max} := \max_{i,j} |A_{ij}|$ to denote the maximum norm. Given a vector $\mathbf{v} = (v_i)_{i=1}^n$, $\|\mathbf{v}\| \equiv \|\mathbf{v}\|_2$ stands for the Euclidean norm. In this paper, we often write an identity matrix as I or 1 without causing any confusion.

Acknowledgements. I want to thank Zongming Ma for bringing this problem to my attention and for valuable suggestions. I also want to thank Edgar Dobriban, David Hong and Yue Sheng for fruitful discussions. I am grateful to the editor, whose feedback has resulted in a significant improvement.

2 The model and main results

2.1 The model

In this paper, we consider the model (1.2). Here X and Y are two independent real matrices of dimensions $p \times n$ and $q \times n$, respectively, where the entries X_{ij}, $1 \leqslant i \leqslant p$, $1 \leqslant j \leqslant n$ and Y_{ij}, $1 \leqslant i \leqslant q$, $1 \leqslant j \leqslant n$ are i.i.d. random variables satisfying that

$$\mathbb{E}X_{11} = \mathbb{E}Y_{11} = 0, \qquad \mathbb{E}|X_{11}|^2 = \mathbb{E}|Y_{11}|^2 = n^{-1}. \tag{2.1}$$

Z is an $r \times n$ matrix that is independent of X and Y, and has i.i.d. entries Z_{ij}, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant n$, satisfying that

$$\mathbb{E}Z_{11} = 0, \qquad \mathbb{E}|Z_{11}|^2 = n^{-1}. \tag{2.2}$$

A and B are $p \times r$ and $q \times r$ deterministic matrices with the following singular value decompositions:

$$A = \sum_{i=1}^{r} a_i \, \mathbf{u}_i^a (\mathbf{v}_i^a)^\top, \quad B = \sum_{i=1}^{r} b_i \, \mathbf{u}_i^b (\mathbf{v}_i^b)^\top, \tag{2.3}$$

where $\{a_i\}$ and $\{b_i\}$ are the singular values, $\{\mathbf{u}_i^a\}$ and $\{\mathbf{u}_i^b\}$ are the left singular vectors, and $\{\mathbf{v}_i^a\}$ and $\{\mathbf{v}_i^b\}$ are the right singular vectors. We assume that for some constant $C > 0$,

$$0 \leqslant a_r \leqslant \cdots \leqslant a_2 \leqslant a_1 \leqslant C, \quad 0 \leqslant b_r \leqslant \cdots \leqslant b_2 \leqslant b_1 \leqslant C. \tag{2.4}$$

For the rest of this paper, we denote that

$$\Sigma_a := \operatorname{diag}(a_1, \cdots, a_r), \quad \Sigma_b := \operatorname{diag}(b_1, \cdots, b_r), \tag{2.5}$$

$$\mathbf{U}_a := \left(\mathbf{u}_1^a, \cdots, \mathbf{u}_r^a\right), \quad \mathbf{V}_a := \left(\mathbf{v}_1^a, \cdots, \mathbf{v}_r^a\right), \quad \mathbf{U}_b := \left(\mathbf{u}_1^b, \cdots, \mathbf{u}_r^b\right), \quad \mathbf{V}_b := \left(\mathbf{v}_1^b, \cdots, \mathbf{v}_r^b\right). \tag{2.6}$$

In this paper, we focus on the high-dimensional setting, that is, there exist constants $\widetilde{c}_1$ and $\widetilde{c}_2$ such that as $n \to \infty$,

$$c_1(n) := \frac{p}{n} \to \widetilde{c}_1, \quad c_2(n) := \frac{q}{n} \to \widetilde{c}_2, \quad \text{with} \quad \widetilde{c}_1 + \widetilde{c}_2 \in (0, 1). \tag{2.7}$$

For simplicity of notations, we will always abbreviate $c_1(n) \equiv c_1$ and $c_2(n) \equiv c_2$ in this paper. Without loss of generality, we can assume that $c_1 \geqslant c_2$. We now summarize the above assumptions for future reference. We will also assume a high moment condition on the entries of X, Y and Z.

Assumption 2.1. *Fix a small constant $\tau > 0$ and large constant $C > 0$.*

(i) *$X = (X_{ij})$ and $Y = (Y_{ij})$ are independent $p \times n$ and $q \times n$ random matrices, whose entries are real i.i.d. random variables satisfying* (2.1) *and the following high moment condition: for any fixed $k \in \mathbb{N}$, there is a constant $\mu_k > 0$ such that*

$$\left(\mathbb{E}|X_{11}|^k\right)^{1/k} \leqslant \mu_k n^{-1/2}, \quad \left(\mathbb{E}|Y_{11}|^k\right)^{1/k} \leqslant \mu_k n^{-1/2}. \tag{2.8}$$

(ii) *$Z = (Z_{ij})$ is an $r \times n$ random matrix independent of X and Y, and its entries are real i.i.d. random variables satisfying* (2.2) *and* (2.8).

(iii) *We assume that*

$$r \leqslant C, \quad \tau \leqslant c_2 \leqslant c_1, \quad c_1 + c_2 \leqslant 1 - \tau. \tag{2.9}$$

(iv) *We consider the model in* (1.2), *where A and B satisfy* (2.3) *and* (2.4).

In this paper, we will use the SCC matrix

$$\mathcal{C}_{\mathcal{X}\mathcal{Y}} := \left(\mathcal{X}\mathcal{X}^\top\right)^{-1/2} \left(\mathcal{X}\mathcal{Y}^\top\right) \left(\mathcal{Y}\mathcal{Y}^\top\right)^{-1} \left(\mathcal{Y}\mathcal{X}^\top\right) \left(\mathcal{X}\mathcal{X}^\top\right)^{-1/2}, \tag{2.10}$$

the null SCC matrix

$$\mathcal{C}_{XY} := S_{xx}^{-1/2} S_{xy} S_{yy}^{-1} S_{yx} S_{xx}^{-1/2}, \tag{2.11}$$

with

$$S_{xx} := XX^\top, \quad S_{yy} := YY^\top, \quad S_{xy} = S_{yx}^\top := XY^\top, \tag{2.12}$$

and the PCC matrix

$$\mathbf{\Sigma}_{\mathcal{X}\mathcal{Y}} := \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1/2}, \tag{2.13}$$

with

$$\Sigma_{xx} := I_p + AA^\top, \quad \Sigma_{yy} := I_q + BB^\top, \quad \Sigma_{xy} = \Sigma_{yx}^\top := AB^\top.$$

We will also use the following SCC, null SCC and PCC matrices:

$$\mathcal{C}_{\mathcal{Y}\mathcal{X}} := \left(\mathcal{Y}\mathcal{Y}^\top\right)^{-1/2}\left(\mathcal{Y}\mathcal{X}^\top\right)\left(\mathcal{X}\mathcal{X}^\top\right)^{-1}\left(\mathcal{X}\mathcal{Y}^\top\right)\left(\mathcal{Y}\mathcal{Y}^\top\right)^{-1/2},$$

$$\mathcal{C}_{YX} = S_{yy}^{-1/2}S_{yx}S_{xx}^{-1}S_{xy}S_{yy}^{-1/2}, \quad \mathbf{\Sigma}_{\mathcal{Y}\mathcal{X}} := \Sigma_{yy}^{-1/2}\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1/2}.$$

Our results can be easily extended to a more general model

$$\mathcal{X} := \mathbf{C}_1^{1/2}X + AZ, \quad \mathcal{Y} := \mathbf{C}_2^{1/2}Y + BZ, \tag{2.14}$$

with non-identity population covariance matrices $\mathbf{C}_1$ and $\mathbf{C}_2$. In fact, it is easy to see that the eigenvalues of the SCC matrix are unchanged under the non-singular transformations $\mathcal{X} \to \mathbf{C}_1^{-1/2}\mathcal{X}$ and $\mathcal{Y} \to \mathbf{C}_2^{-1/2}\mathcal{Y}$, which reduce (2.14) to the model (1.2) with A and B replaced by $\mathbf{C}_1^{-1/2}A$ and $\mathbf{C}_2^{-1/2}B$.

2.2 The main results

We denote the eigenvalues of the null SCC matrix $\mathcal{C}_{YX}$ by $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_q \geqslant 0$. It is easy to see that $\mathcal{C}_{XY}$ shares the same eigenvalues with $\mathcal{C}_{YX}$, besides the $p-q$ more trivial zero eigenvalues $\lambda_{q+1} = \cdots = \lambda_p = 0$. We denote the ESD of $\mathcal{C}_{YX}$ by

$$F_n(x) := \frac{1}{q}\sum_{i=1}^{q}\mathbf{1}_{\lambda_i \leqslant x}.$$

It has been proved [37, 42] that, almost surely, F_n converges weakly to a deterministic probability distribution $F(x)$ with density

$$f(x) = \frac{1}{2\pi c_2}\frac{\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{x(1-x)}, \quad \lambda_- \leqslant x \leqslant \lambda_+, \tag{2.15}$$

where

$$\lambda_\pm := \left(\sqrt{c_1(1-c_2)} \pm \sqrt{c_2(1-c_1)}\right)^2. \tag{2.16}$$

For the model (1.2) with finite rank correlations, we denote the eigenvalues of $\mathcal{C}_{\mathcal{Y}\mathcal{X}}$ by $\widetilde{\lambda}_1 \geqslant \widetilde{\lambda}_2 \geqslant \cdots \geqslant \widetilde{\lambda}_q \geqslant 0$, and the eigenvalues of $\mathbf{\Sigma}_{\mathcal{Y}\mathcal{X}}$ by

$$t_1 \geqslant t_2 \geqslant \cdots \geqslant t_r \geqslant t_{r+1} = \cdots = t_q = 0. \tag{2.17}$$

Recall the threshold t_c defined in (1.3) for BBP transition. If the entries of X and Y are i.i.d. Gaussian, then it was proved in [4] that the following phenomena occur for any $1 \leqslant i \leqslant r$: if $t_i \leqslant t_c$, then $\widetilde{\lambda}_i - \lambda_+ \to 0$ almost surely; if $t_i > t_c$, then $\widetilde{\lambda}_i - \theta_i \to 0$ almost surely, where

$$\theta_i := t_i\left(1 - c_1 + c_1 t_i^{-1}\right)\left(1 - c_2 + c_2 t_i^{-1}\right). \tag{2.18}$$

Moreover, the limiting distributions were also identified in [4] for the Gaussian case: if $t_i < t_c$, $n^{2/3}(\widetilde{\lambda}_i - \lambda_+)$ converges to the Tracy-Widom law; if $t_i > t_c$, $\sqrt{n}(\widetilde{\lambda}_i - \theta_i)$ converges to a centered normal distribution. The main purpose of this paper is to extend the CLT of the outliers to the setting in Section 2.1, assuming only the moment conditions in (2.8) (or the weaker ones in (2.34) below).

In [4], it was assumed that that the t_i's are either well-separate or exactly degenerate. In this paper, however, we consider the most general setting which allows for near-degenerate outliers. For this purpose, we first introduce some new notations following [27]. For any $r \times r$ matrix $\mathcal{A} = (A_{ij})$ and a subset of indices $\pi \subset \{1, \cdots, r\}$, we define the $|\pi| \times |\pi|$ matrix $\mathcal{A}_{[\![\pi]\!]}$ by

$$\mathcal{A}_{[\![\pi]\!]} := (A_{ij})_{i,j \in \pi}. \tag{2.19}$$

We denote the eigenvalues of $\mathcal{A}_{[\![\pi]\!]}$ in the descending order by

$$\mu_1\left(\mathcal{A}_{[\![\pi]\!]}\right) \geqslant \cdots \geqslant \mu_{|\pi|}\left(\mathcal{A}_{[\![\pi]\!]}\right). \tag{2.20}$$

We will group the near-degenerate t_i's according to the following definition.

Definition 2.2. *Fix two small constants $\delta_l, \delta > 0$. For $l \in \{1, \cdots, r\}$ satisfying*

$$t_c + \delta_l \leqslant t_l \leqslant 1 - \delta_l, \tag{2.21}$$

we define the subset $\gamma(l) \ni l$ as the smallest *subset of $\{1, \cdots, r\}$ such that the following property holds: if $i, j \in \{1, \cdots, r\}$ satisfy $t_i > t_c$ and $|t_i - t_j| \leqslant n^{-1/2+\delta}$, then either $i, j \in \gamma(l)$ or $i, j \notin \gamma(l)$.*

The set $\gamma(l)$ in this definition can be constructed by successively choosing $i \in \{1, \cdots, r\}$ such that t_i is away from $\gamma(l)$ by a distance less than $n^{-1/2+\delta}$, and then adding i to $\gamma(l)$. Since the number of such indices is at most r, we have that $|t_i - t_l| \leqslant rn^{-1/2+\delta}$ for any $i \in \gamma(l)$.

Before stating our main result, we first define the reference matrix, which describes the limiting distribution of a group of near-degenerate outliers. Recalling (2.5) and (2.6), we abbreviate that

$$\widehat{\Sigma}_a := \frac{\Sigma_a}{(I_r + \Sigma_a^2)^{1/2}}, \quad \widehat{\Sigma}_b := \frac{\Sigma_b}{(I_r + \Sigma_b^2)^{1/2}}, \quad \mathcal{M}_r := \mathbf{V}_a^\top \mathbf{V}_b. \tag{2.22}$$

Since $\widehat{\Sigma}_a \mathcal{M}_r \widehat{\Sigma}_b^2 \mathcal{M}_r^\top \widehat{\Sigma}_a$ is the PCC matrix in the basis of $\{\mathbf{u}_i^a\}$, we have the singular value decomposition

$$\widehat{\Sigma}_a \mathcal{M}_r \widehat{\Sigma}_b = \mathcal{O}\,\mathrm{diag}(\sqrt{t_1}, \cdots, \sqrt{t_r})\widetilde{\mathcal{O}}^\top, \tag{2.23}$$

for some $r \times r$ orthogonal matrices $\mathcal{O}$ and $\widetilde{\mathcal{O}}$. Fix any $1 \leqslant l \leqslant r$ such that (2.21) holds. We define

$$\begin{aligned}\mathcal{W}_{k,ij} \equiv \mathcal{W}_{k,ji} := {}& t_l \left[\mathbf{V}_a \widehat{\Sigma}_a \mathcal{O}\right]_{ki} \left[\mathbf{V}_a \widehat{\Sigma}_a \mathcal{O}\right]_{kj} + t_l \left[\mathbf{V}_b \widehat{\Sigma}_b \widetilde{\mathcal{O}}\right]_{ki} \left[\mathbf{V}_b \widehat{\Sigma}_b \widetilde{\mathcal{O}}\right]_{kj} \\ & - \sqrt{t_l} \left[\mathbf{V}_a \widehat{\Sigma}_a \mathcal{O}\right]_{ki} \left[\mathbf{V}_b \widehat{\Sigma}_b \widetilde{\mathcal{O}}\right]_{kj} - \sqrt{t_l} \left[\mathbf{V}_b \widehat{\Sigma}_b \widetilde{\mathcal{O}}\right]_{ki} \left[\mathbf{V}_a \widehat{\Sigma}_a \mathcal{O}\right]_{kj},\end{aligned} \tag{2.24}$$

and

$$\mathcal{U} := \mathbf{U}_a \left(I_r - \widehat{\Sigma}_a^2\right)^{1/2} \mathcal{O} = \mathbf{U}_a \frac{1}{(I_r + \Sigma_a^2)^{1/2}} \mathcal{O}, \quad \mathcal{V} := \mathbf{U}_b \left(I_r - \widehat{\Sigma}_b^2\right)^{1/2} \widetilde{\mathcal{O}} = \mathbf{U}_b \frac{1}{(1 + \Sigma_b^2)^{1/2}} \widetilde{\mathcal{O}}. \tag{2.25}$$

Then we define the following covariance function for $(i,j), (i',j') \in \gamma(l) \times \gamma(l)$:

$$\begin{aligned}C_{ij,i'j'}(t_l) := {}& \frac{(1-t_l)^2 t_l^2}{t_l^2 - t_c^2}\left(2t_l + \frac{c_1}{1-c_1} + \frac{c_2}{1-c_2}\right)(\delta_{ii'}\delta_{jj'} + \delta_{ij'}\delta_{ji'}) \\ & + t_l^2(\mu_x^{(4)} - 3)\sum_k \mathcal{U}_{ki}\mathcal{U}_{ki'}\mathcal{U}_{kj}\mathcal{U}_{kj'} + t_l^2(\mu_y^{(4)} - 3)\sum_k \mathcal{V}_{ki}\mathcal{V}_{ki'}\mathcal{V}_{kj}\mathcal{V}_{kj'} + (\mu_z^{(4)} - 3)\sum_k \mathcal{W}_{k,ij}\mathcal{W}_{k,i'j'},\end{aligned} \tag{2.26}$$

where we have used the notations

$$\mu_x^{(4)} := n^2 \mathbb{E} X_{11}^4, \quad \mu_y^{(4)} := n^2 \mathbb{E} Y_{11}^4, \quad \mu_z^{(4)} := n^2 \mathbb{E} Z_{11}^4. \tag{2.27}$$

Finally, we define the notation

$$a(t_l) := \frac{(1-c_1)(1-c_2)}{t_l^2}(t_l^2 - t_c^2). \tag{2.28}$$

Now we are ready to state the main results.

Theorem 2.3. *Fix any $1 \leqslant l \leqslant r$. Suppose Assumption 2.1 and* (2.21) *hold. Define the vector of rescaled eigenvalues $\boldsymbol{\zeta} = (\zeta_i)_{i\in\gamma(l)}$, where $\zeta_i := n^{1/2}(\widetilde{\lambda}_i - \theta_i)$ for θ_i defined in* (2.18)*. Let $\boldsymbol{\xi} = (\xi_i)_{i\in\gamma(l)}$ be the vector of ordered eigenvalues of the random $|\gamma(l)| \times |\gamma(l)|$ matrix*

$$a(t_l)n^{1/2}\left[\mathrm{diag}(t_1, \cdots, t_r) - t_l\right]_{[\![\gamma(l)]\!]} + a(t_l)\Upsilon_l, \tag{2.29}$$

where Υ_l is a $|\gamma(l)|\times|\gamma(l)|$ symmetric Gaussian random matrix, whose entries have mean zero and covariances

$$\mathbb{E}(\Upsilon_l)_{ij}(\Upsilon_l)_{i'j'} = C_{ij,i'j'}(t_l), \quad \textit{for} \quad (i,j),(i',j') \in \gamma(l) \times \gamma(l). \tag{2.30}$$

Then for any bounded continuous function f, we have that

$$\lim_n \left[\mathbb{E}f(\boldsymbol{\zeta}) - \mathbb{E}f(\boldsymbol{\xi})\right] = 0. \tag{2.31}$$

By Portmanteau theorem, (2.31) means that the group of rescaled eigenvalues $\boldsymbol{\zeta}$ has the same joint asymptotic distribution as the eigenvalues of a Gaussian random matrix with covariances given by (2.30) (up to a factor $a^2(t_l)$). Now we apply our result to the special case where the entries of X, Y and Z are i.i.d. Gaussian random variables, and $t_i = t_l$ for all $i \in \gamma(l)$. In this case, the last three terms in (2.26) and the first term in (2.29) vanish. Then by Theorem 2.3, we get that $\boldsymbol{\zeta}$ converges weakly to the ordered eigenvalues of a GOE $G = (g_{ij})$, with independent Gaussian entries

$$g_{ij} = g_{ji} \sim \mathcal{N}(0, (1+\delta_{ij})c_g(t_l)), \tag{2.32}$$

where

$$c_g(t_l) := \frac{(1-c_1)^2(1-c_2)^2(1-t_l)^2(t_l^2-t_c^2)}{t_l^2}\left(2t_l + \frac{c_1}{1-c_1} + \frac{c_2}{1-c_2}\right). \tag{2.33}$$

This is in accordance with [4, Theorem 1.9].

We can relax the moment assumption (2.8) if we assume that the population CCC are either well-separated or exactly degenerate.

Theorem 2.4. *Fix any $1 \leqslant l \leqslant r$. Suppose Assumption 2.1 holds except that* (2.8) *is replaced with the following moment assumption: there exist constants $c_0, C_0 > 0$ such that*

$$\mathbb{E}|\sqrt{n}X_{11}|^{8+c_0} \leqslant C_0, \quad \mathbb{E}|\sqrt{n}Y_{11}|^{8+c_0} \leqslant C_0, \quad \mathbb{E}|\sqrt{n}Z_{11}|^{4+c_0} \leqslant C_0. \tag{2.34}$$

Suppose there exists a constant $\delta_l > 0$ such that (2.21) *holds, and*

$$t_i = t_l \quad \textit{for} \quad i \in \gamma(l), \qquad \textit{and} \qquad |t_i - t_l| \geqslant \delta_l \quad \textit{for} \quad i \notin \gamma(l). \tag{2.35}$$

Then the equation (2.31) *holds for $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ defined in Theorem 2.3.*

On the other hand, the limiting Tracy-Widom distribution of the extreme non-outlier eigenvalues has been proved under a fourth moment tail assumption in [31].

Theorem 2.5 (Theorem 2.14 of [31])**.** *Suppose Assumption 2.1 (iii)-(iv) hold. Assume that $x_{ij} = n^{-1/2}\widehat{x}_{ij}$, $y_{ij} = n^{-1/2}\widehat{y}_{ij}$ and $z_{ij} = n^{-1/2}\widehat{z}_{ij}$, where $\{\widehat{x}_{ij}\}$, $\{\widehat{y}_{ij}\}$ and $\{\widehat{z}_{ij}\}$ are three independent families of real i.i.d. random variables with mean zero and variance one. Moreover, we assume the fourth moment tail condition*

$$\lim_{t\to\infty} t^4 \left[\mathbb{P}\left(|\widehat{x}_{11}| \geqslant t\right) + \mathbb{P}\left(|\widehat{y}_{11}| \geqslant t\right)\right] = 0. \tag{2.36}$$

Assume that for a fixed $0 \leqslant r_+ \leqslant r$, the eigenvalues of $\boldsymbol{\Sigma}_{\mathcal{XY}}$ satisfy

$$\liminf_n t_{r_+} > t_c > \limsup_n t_{r_++1}. \tag{2.37}$$

Then we have that for any fixed $k \in \mathbb{N}$,

$$\lim_{n\to\infty} \mathbb{P}\left[\left(n^{2/3}\frac{\widetilde{\lambda}_{r_++i}-\lambda_+}{c_{TW}} \leqslant s_i\right)_{i=1}^{k}\right] = \lim_{n\to\infty} \mathbb{P}^{GOE}\left[\left(n^{2/3}(\lambda_i - 2) \leqslant s_i\right)_{i=1}^{k}\right], \tag{2.38}$$

for all $(s_1, s_2, \ldots, s_k) \in \mathbb{R}^k$, where

$$c_{TW} := \left[\frac{\lambda_+^2(1-\lambda_+)^2}{\sqrt{c_1c_2(1-c_1)(1-c_2)}}\right]^{1/3},$$

and $\mathbb{P}^{GOE}$ stands for the law of GOE (Gaussian orthogonal ensemble).

Taking $k = 1$ in (2.38) gives that $n^{2/3}(\widetilde{\lambda}_{r_++1} - \lambda_+)/c_{TW}$ converges weakly to the type-1 Tracy-Widom distribution [35, 36]. Furthermore, the joint distribution of the largest k eigenvalues of GOE can be written in terms of the Airy kernel [18]. Hence (2.38) gives a complete description of the asymptotic finite-dimensional joint distributions of the extreme non-outlier eigenvalues of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$. Combining Theorems 2.3, 2.4 and 2.5, we complete the story of BBP transition for high-dimensional CCA with finite rank correlations.

Before the end of this section, we verify Theorem 2.3 with some numerical simulations. In particular, we will show that the last three terms in (2.26) are necessary to match the variance of the simulated sample CCC. For the simulations, we take the entries of X, Y and Z to be i.i.d. Rademacher random variables (with an extra scaling $n^{-1/2}$). In this setting, we have $\mu_x^{(4)} = \mu_y^{(4)} = \mu_z^{(4)} = 1$. Moreover, we take $n = 2000$ and $c_1 = c_2 = 0.2$, i.e. $p = q = 400$, which gives $t_c = 0.25$ by (1.3). We consider the rank-one case with $r = 1$, and take the matrices A and B as $A = 2\mathbf{u}^a$ and $B = 2\mathbf{u}^b$, which gives a supercritical spike $t_1 = 0.64$. We consider the following two scenarios for the unit vectors $\mathbf{u}^a$ and $\mathbf{u}^b$.

Scenario (a): $\mathbf{u}^a$ and $\mathbf{u}^b$ are standard unit vectors along the first coordinate axis in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. In this case, the limiting variance of $\zeta_1 = n^{1/2}(\widetilde{\lambda}_1 - \theta_1)$ is given by $\sigma_a^2 := a^2(t_1)C_{11,11}(t_1)$, where by (2.26) we have that

$$\begin{aligned} C_{11,11}(t_1) = {} & 2\frac{(1-t_1)^2t_1^2}{t_1^2-t_c^2}\left(2t_1 + \frac{c_1}{1-c_1} + \frac{c_2}{1-c_2}\right) - 2t_1^2\left[\frac{1}{(1+a^2)^2} + \frac{1}{(1+b^2)^2}\right] \\ & - 2\left[t_1\frac{a^2}{1+a^2} + t_1\frac{b^2}{1+b^2} - 2\sqrt{t_1}\frac{ab}{(1+a^2)^{1/2}(1+b^2)^{1/2}}\right]^2. \end{aligned}$$

Scenario (b): $\mathbf{u}^a$ and $\mathbf{u}^b$ are random unit vectors on the unit spheres $\mathbb{S}^p$ and $\mathbb{S}^q$, respectively. Then we have $\|\mathbf{u}^a\|_\infty \leqslant n^{-1/2+\varepsilon}$ and $\|\mathbf{u}^b\|_\infty \leqslant n^{-1/2+\varepsilon}$ with probability $1 - \mathrm{o}(1)$ for any constant $\varepsilon > 0$, with which we can easily check that the $\mathcal{U}$ and $\mathcal{V}$ terms in (2.26) are both of order $\mathrm{O}(n^{-1+2\varepsilon})$ with probability $1 - \mathrm{o}(1)$. Hence the limiting variance of ζ_1 is given by $\sigma_b^2 + \mathrm{o}(1)$ with probability $1 - \mathrm{o}(1)$, where

$$\begin{aligned} \sigma_b^2 := {} & 2a^2(t_1)\frac{(1-t_1)^2t_1^2}{t_1^2-t_c^2}\left(2t_1 + \frac{c_1}{1-c_1} + \frac{c_2}{1-c_2}\right) \\ & - 2a^2(t_1)\left[t_1\frac{a^2}{1+a^2} + t_1\frac{b^2}{1+b^2} - 2\sqrt{t_1}\frac{ab}{(1+a^2)^{1/2}(1+b^2)^{1/2}}\right]^2. \end{aligned}$$

In Figure 1, we report the simulation results based on 5000 replications. We find that the histograms match our result in Theorem 2.3 pretty well. Furthermore, it is not surprising that the predication (2.32) in the Gaussian setting deviates from the simulations a little bit, which shows that the last three terms in (2.26) are necessary for the non-Gaussian setting.

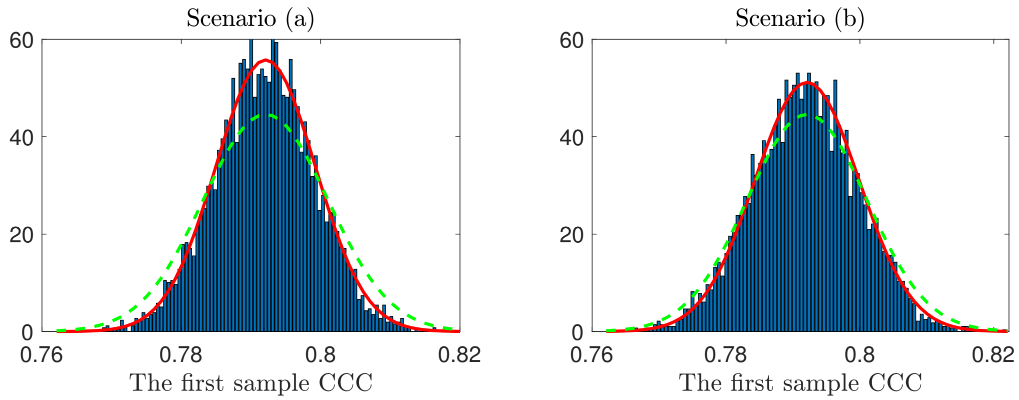

Figure 1: The histograms give the simulated first sample CCC based on 5000 replications. The red solid curves give the probability density functions (PDF) of the normal distributions $\mathcal{N}(\theta_1, \sigma_a^2/n)$ and $\mathcal{N}(\theta_1, \sigma_b^2/n)$ in scenarios (a) and (b), respectively. The green dashed curves represent the PDF of the normal distribution $\mathcal{N}(\theta_1, 2c_g(t_1)/n)$, where $c_g(t_1)$ is defined in (2.33).

3 Linearization method and resolvents

In this section, we introduce the linearization method, define the concept of *resolvent*, and show that the study of the limiting distributions of the outliers can be reduced to proving the CLT for a linear functional of the resolvent. Before that, we first recall some (almost) sharp convergence estimates on the sample CCCs that have been proved in [31, 40]. They will serve as important a priori estimates for our proof.

3.1 Convergence of the sample CCCs

To simplify the notations, it is helpful to use the following notion of stochastic domination introduced in [13]. It simplifies the presentation of the results and their proofs by systematizing statements of the form "ξ is bounded by ζ with high probability up to a small power of n".

Definition 3.1 (Stochastic domination and high probability event). *(i) Let*

$$\xi = \left(\xi^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}\right), \quad \zeta = \left(\zeta^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}\right)$$

be two families of nonnegative random variables, where $U^{(n)}$ is a possibly n-dependent parameter set. We say ξ is stochastically dominated by ζ, uniformly in u, if for any small constant $\varepsilon > 0$ and large constant $D > 0$, we have that

$$\sup_{u \in U^{(n)}} \mathbb{P}\left[\xi^{(n)}(u) > n^{\varepsilon} \zeta^{(n)}(u)\right] \leqslant n^{-D}$$

for large enough $n \geqslant n_0(\varepsilon, D)$, and we will use the notation $\xi \prec \zeta$ to denote it. If a family of complex random variables ξ satisfy $|\xi| \prec \zeta$, then we will also write $\xi \prec \zeta$ or $\xi = \mathrm{O}_\prec(\zeta)$.

(ii) We extend $\mathrm{O}_\prec(\cdot)$ to matrices in the operator norm sense as follows. Let A be a family of random matrices and ζ be a family of nonnegative random variables. Then $A = \mathrm{O}_\prec(\zeta)$ means that $\|A\| \prec \zeta$.

(iii) As a convention, for two deterministic *nonnegative quantities ξ and ζ, we write $\xi \prec \zeta$ if and only if $\xi \leqslant n^\tau \zeta$ for any constant $\tau > 0$.*

(iv) We say an event Ξ holds with high probability (w.h.p.) if for any constant $D > 0$, $\mathbb{P}(\Xi) \geqslant 1 - n^{-D}$ for large enough n. Moreover, we say Ξ holds with high probability on an event Ω if for any constant $D > 0$, $\mathbb{P}(\Omega \backslash \Xi) \leqslant n^{-D}$ for large enough n.

The following lemma collects some basic properties of stochastic domination $\prec$, which will be used tacitly in the proof.

Lemma 3.2 (Lemma 3.2 in [8])**.** *Let ξ and ζ be two families of nonnegative random variables, U and V be two parameter sets, and $C > 0$ be a large constant.*

(i) Suppose that $\xi(u, v) \prec \zeta(u, v)$ uniformly in $u \in U$ and $v \in V$. If $|V| \leqslant n^C$, then $\sum_{v \in V} \xi(u, v) \prec \sum_{v \in V} \zeta(u, v)$ uniformly in $u \in U$.

(ii) If $\xi_1(u) \prec \zeta_1(u)$ and $\xi_2(u) \prec \zeta_2(u)$ uniformly in $u \in U$, then $\xi_1(u)\xi_2(u) \prec \zeta_1(u)\zeta_2(u)$ uniformly in $u \in U$.

(iii) Suppose that $\Psi(u) \geqslant n^{-C}$ is deterministic and $\xi(u)$ satisfies $\mathbb{E}\xi(u)^2 \leqslant n^C$ for all $u \in U$. Then if $\xi(u) \prec \Psi(u)$ uniformly in $u \in U$, we have that $\mathbb{E}\xi(u) \prec \Psi(u)$ uniformly in $u \in U$.

The following large deviation bounds on the outliers of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ were proved in [31].

Lemma 3.3 (Theorem 2.9 of [31])**.** *Suppose Assumption 2.1 holds. If $t_i \geqslant t_c + n^{-1/3}$, then we have that*

$$|\widetilde{\lambda}_i - \theta_i| \prec n^{-1/2} |t_i - t_c|^{1/2}. \tag{3.1}$$

On the other hand, for any $i = \mathrm{O}(1)$ with $t_i < t_c + n^{-1/3}$, we have that

$$|\widetilde{\lambda}_i - \lambda_+| \prec n^{-2/3}. \tag{3.2}$$

The quantiles of the density (2.15) correspond to the classical locations of the eigenvalues of $\mathcal{C}_{YX}$.

Definition 3.4. *The classical location γ_j of the j-th eigenvalue of $\mathcal{C}_{YX}$ is defined as*

$$\gamma_j := \sup_x \left\{ \int_x^{+\infty} f(x) \mathrm{d}x > \frac{j-1}{q} \right\}, \tag{3.3}$$

where f is defined in (2.15). Note that we have $\gamma_1 = \lambda_+$ and $\lambda_+ - \gamma_j \sim (j/n)^{2/3}$ for $j > 1$.

In [40], we have proved the following eigenvalue rigidity estimate for $\mathcal{C}_{YX}$.

Theorem 3.5 (Theorem 2.5 of [40])**.** *Suppose Assumption 2.1 holds. The eigenvalues of the null SCC matrix $\mathcal{C}_{YX}$ satisfy the following eigenvalue rigidity estimate:*

$$|\lambda_i - \gamma_i| \prec i^{-1/3} n^{-2/3}, \quad 1 \leqslant i \leqslant (1-\delta)q, \tag{3.4}$$

where $\delta > 0$ is any small constant.

3.2 Resolvents and local laws

One main tool of our proof is the self-adjoint linearization trick developed in [31, 40]: a $\lambda \in (0, 1)$ is an eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ if and only if the following equation holds:

$$\det \begin{bmatrix} 0 & \begin{pmatrix} \mathcal{X} & 0 \\ 0 & \mathcal{Y} \end{pmatrix} \\ \begin{pmatrix} \mathcal{X}^\top & 0 \\ 0 & \mathcal{Y}^\top \end{pmatrix} & \begin{pmatrix} \lambda I_n & \lambda^{1/2} I_n \\ \lambda^{1/2} I_n & \lambda I_n \end{pmatrix}^{-1} \end{bmatrix} = 0. \tag{3.5}$$

Inspired by this equation, we define the following $(p+q+2n)\times(p+q+2n)$ self-adjoint block matrix

$$H(\lambda)\equiv H(X,Y,\lambda):=\begin{bmatrix} 0 & \begin{pmatrix} X & 0\\ 0 & Y\end{pmatrix}\\ \begin{pmatrix} X^\top & 0\\ 0 & Y^\top\end{pmatrix} & \begin{pmatrix} \lambda I_n & \lambda^{1/2}I_n\\ \lambda^{1/2}I_n & \lambda I_n\end{pmatrix}^{-1}\end{bmatrix}, \tag{3.6}$$

and call its inverse the *resolvent*:

$$G(\lambda)\equiv G(X,Y,\lambda):=[H(\lambda)]^{-1}. \tag{3.7}$$

We can also extend the argument λ to $z\in\mathbb{C}_+:=\{z\in\mathbb{C}:\operatorname{Im} z>0\}$ with $z^{1/2}$ being the branch with positive imaginary part. Similar to equation (3.5), it is not hard to see that λ is not an eigenvalue of the null SCC matrix if and only if $\det[H(\lambda)]\neq 0$. In this case, using (1.2), (2.3), (2.5) and (2.6), we can rewrite (3.5) as

$$\begin{aligned} 0&=\det\left[1+\begin{pmatrix}\mathbf{U} & 0\\ 0 & \mathbf{E}\end{pmatrix}\begin{pmatrix}0 & \mathcal{D}\\ \mathcal{D} & 0\end{pmatrix}\begin{pmatrix}\mathbf{U}^\top & 0\\ 0 & \mathbf{E}^\top\end{pmatrix}G(\lambda)\right]\\ &=\det\left[1+\begin{pmatrix}0 & \mathcal{D}\\ \mathcal{D} & 0\end{pmatrix}\begin{pmatrix}\mathbf{U}^\top & 0\\ 0 & \mathbf{E}^\top\end{pmatrix}G(\lambda)\begin{pmatrix}\mathbf{U} & 0\\ 0 & \mathbf{E}\end{pmatrix}\right], \end{aligned} \tag{3.8}$$

where we used the identity $\det(1+M_1M_2)=\det(1+M_2M_1)$ for any two matrices M_1 and M_2 of conformable dimensions. Here $\mathcal{D}$, $\mathbf{U}$ and $\mathbf{E}$ are $2r\times 2r$, $(p+q)\times 2r$ and $2n\times 2r$ matrices defined as

$$\mathcal{D}:=\begin{pmatrix}\Sigma_a & 0\\ 0 & \Sigma_b\end{pmatrix},\quad \mathbf{U}:=\begin{pmatrix}\mathbf{U}_a & 0\\ 0 & \mathbf{U}_b\end{pmatrix},\quad \mathbf{E}:=\begin{pmatrix}Z^\top\mathbf{V}_a & 0\\ 0 & Z^\top\mathbf{V}_b\end{pmatrix}.$$

In [40], we have proved an (almost) sharp large deviation estimate on the $4r\times 4r$ matrix

$$\begin{pmatrix}\mathbf{U}^\top & 0\\ 0 & \mathbf{E}^\top\end{pmatrix}G(\lambda)\begin{pmatrix}\mathbf{U} & 0\\ 0 & \mathbf{E}\end{pmatrix},$$

called the *anisotropic local law*. Before stating it, we first introduce some notations.

Definition 3.6 (Index sets). *For simplicity of notations, we define the index sets*

$$\mathcal{I}_1:=\{1,\cdots,p\},\quad \mathcal{I}_2:=\{p+1,\cdots,p+q\},$$
$$\mathcal{I}_3:=\{p+q+1,\cdots,p+q+n\},\quad \mathcal{I}_4:=\{p+q+n+1,\cdots,p+q+2n\}.$$

We will consistently use latin letters $i,j\in\mathcal{I}_1\cup\mathcal{I}_2$ and greek letters $\mu,\nu\in\mathcal{I}_3\cup\mathcal{I}_4$. Moreover, we will use the notations $\mathfrak{a},\mathfrak{b}\in\mathcal{I}:=\cup_{i=1}^4\mathcal{I}_i$.

Denote the averaged partial traces of the resolvent by

$$m_\alpha(z):=\frac{1}{n}\sum_{\mathfrak{a}\in\mathcal{I}_\alpha}G_{\mathfrak{a}\mathfrak{a}}(z),\quad \alpha=1,2,3,4. \tag{3.9}$$

In [40], we have shown that these partial traces converge to deterministic limits given by

$$m_{1c}(z)=\frac{-z+c_1+c_2+\sqrt{(z-\lambda_-)(z-\lambda_+)}}{2(1-c_1)z(1-z)}-\frac{c_1}{(1-c_1)z}, \tag{3.10}$$

$$m_{2c}(z)=\frac{-z+c_1+c_2+\sqrt{(z-\lambda_-)(z-\lambda_+)}}{2(1-c_2)z(1-z)}-\frac{c_2}{(1-c_2)z}, \tag{3.11}$$

$$m_{3c}(z)=\frac{1}{2}\left[(1-2c_1)z+c_1-c_2+\sqrt{(z-\lambda_-)(z-\lambda_+)}\right], \tag{3.12}$$
$$m_{4c}(z)=\frac{1}{2}\left[(1-2c_2)z+c_2-c_1+\sqrt{(z-\lambda_-)(z-\lambda_+)}\right], \tag{3.13}$$

where $\lambda_\pm$ is defined in (2.16). In [40], we also verified the following equations for $m_{\alpha c}$:

$$m_{1c}=-\frac{c_1}{m_{3c}},\quad m_{2c}=-\frac{c_2}{m_{4c}},\quad m_{3c}(z)-m_{4c}(z)=(1-z)(c_1-c_2), \tag{3.14}$$
$$m_{3c}(z)=\frac{1-(z-1)m_{2c}(z)}{z^{-1}-[m_{1c}(z)+m_{2c}(z)]+(z-1)m_{1c}(z)m_{2c}(z)}, \tag{3.15}$$
$$m_{4c}(z)=\frac{1-(z-1)m_{1c}(z)}{z^{-1}-[m_{1c}(z)+m_{2c}(z)]+(z-1)m_{1c}(z)m_{2c}(z)}. \tag{3.16}$$

One can also check them through direct calculations with (3.10)–(3.13). We also define the function

$$\begin{aligned} h(z) &:= \frac{z^{-1/2}m_{3c}(z)}{1+(1-z)m_{2c}(z)}=\frac{z^{-1/2}m_{4c}(z)}{1+(1-z)m_{1c}(z)} \\ &= \frac{z^{1/2}}{2}\left[-z+(2-c_1-c_2)+\sqrt{(z-\lambda_-)(z-\lambda_+)}\right]. \end{aligned} \tag{3.17}$$

Now with the above definitions, we define the matrix limit of $G(z)$ as

$$\Pi(z):=\begin{bmatrix}\begin{pmatrix}c_1^{-1}m_{1c}(z)I_p & 0\\ 0 & c_2^{-1}m_{2c}(z)I_q\end{pmatrix} & 0\\ 0 & \begin{pmatrix}m_{3c}(z)I_n & h(z)I_n\\ h(z)I_n & m_{4c}(z)I_n\end{pmatrix}\end{bmatrix}. \tag{3.18}$$

Using (3.14)–(3.17), it is easy to check that

$$\Pi=\begin{bmatrix}\begin{pmatrix}-m_{3c}I_p & 0\\ 0 & -m_{4c}I_q\end{pmatrix} & 0\\ 0 & \begin{pmatrix}zI_n & z^{1/2}I_n\\ z^{1/2}I_n & zI_n\end{pmatrix}^{-1}-\begin{pmatrix}m_{1c}I_n & 0\\ 0 & m_{2c}I_n\end{pmatrix}\end{bmatrix}^{-1}. \tag{3.19}$$

We define two different spectral domains of z for our statements of the local laws.

Definition 3.7. *For any constant $\varepsilon>0$, we define a spectral domain around the bulk spectrum $[\lambda_-,\lambda_+]$ as*

$$S(\varepsilon):=\left\{z=E+\mathrm{i}\eta:\varepsilon\leqslant E\leqslant 1-\varepsilon,n^{-1+\varepsilon}\leqslant\eta\leqslant\varepsilon^{-1}\right\}, \tag{3.20}$$

and a spectral domain outside the bulk spectrum as

$$S_{out}(\varepsilon):=\left\{z=E+\mathrm{i}\eta:\lambda_++n^{-2/3+\varepsilon}\leqslant E\leqslant 1-\varepsilon,0\leqslant\eta\leqslant\varepsilon^{-1}\right\}. \tag{3.21}$$

The following theorem gives the anisotropic local laws for the resolvent $G(z)$ on the above two spectral domains. It is one of the most important tools for our proof.

Theorem 3.8 (Anisotropic local laws)**.** *Suppose Assumption 2.1 holds. For any fixed $\varepsilon>0$ and deterministic unit vectors $\mathbf{u},\mathbf{v}\in\mathbb{C}^{\mathcal{I}}$, the following anisotropic local laws hold.*

(i) **(Theorem 2.11 of [40])**. *For any $z = E + \mathrm{i}\eta \in S(\varepsilon)$, we have that*

$$|\langle \mathbf{u}, G(z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec \Psi(z) + (n\eta)^{-1}, \tag{3.22}$$

where the inner product is defined as $\langle \mathbf{v}, \mathbf{w}\rangle := \mathbf{v}^* \mathbf{w}$ with $\mathbf{v}^*$ denoting the conjugate transpose, and $\Psi(z)$ is the control parameter defined as

$$\Psi(z) := \sqrt{\frac{\operatorname{Im} m_{3c}(z)}{n\eta}}. \tag{3.23}$$

(ii) **(Theorem 3.10 of [31])**. *For any $z = E + \mathrm{i}\eta \in S_{out}(\varepsilon)$, we have that*

$$|\langle \mathbf{u}, G(z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec n^{-1/2}(\kappa + \eta)^{-1/4}, \tag{3.24}$$

where κ is the distance to the right edge along the real axis:

$$\kappa \equiv \kappa_z := |E - \lambda_+|. \tag{3.25}$$

The above estimates (3.22) *and* (3.24) *hold uniformly in the spectral parameter z. Moreover, for these estimates to hold, it is not necessary to assume that the entries of X, Y and Z are identically distributed, that is, only independence and moment conditions are needed.*

The averaged partial traces in (3.9) satisfy the stronger averaged local laws.

Theorem 3.9 (Averaged local law, Theorem 2.12 of [40])**.** *Suppose Assumption 2.1 holds. For any fixed $\varepsilon > 0$, we have that*

$$|m_\alpha(z) - m_{\alpha c}(z)| \prec (n\eta)^{-1}, \quad \alpha = 1, 2, 3, 4, \tag{3.26}$$

uniformly in $z \in S(\varepsilon)$. Moreover, outside of the spectrum we have the stronger estimate

$$|m_\alpha(z) - m_{\alpha c}(z)| \prec \frac{1}{n(\kappa + \eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa + \eta}}, \quad \alpha = 1, 2, 3, 4, \tag{3.27}$$

uniformly in $z \in S(\varepsilon) \cap S_{out}(\varepsilon)$.

3.3 Reduction to the law of resolvent

In this subsection, we reduce the study of the limiting law of $\boldsymbol{\zeta}$ in Theorem 2.3 to the study of the limiting law of a linear functional of the resolvent $G(z)$. Without loss of generality, we assume a slightly stronger condition than (2.4) so that A and B are both of rank r:

$$0 < a_r \leqslant \cdots \leqslant a_2 \leqslant a_1 \leqslant C, \quad 0 < b_r \leqslant \cdots \leqslant b_2 \leqslant b_1 \leqslant C. \tag{3.28}$$

This can be achieved by adding a small $0 < \varepsilon_n < e^{-n}$ to each zero a_i or b_i. Since the proof does not depend on the lower bounds of a_r and b_r, we can easily extend it to the case with zero a_i's or b_i's by taking $\varepsilon_n \to 0$.

Recall that if $\lambda \in (0, 1)$ is not in the spectrum of $\mathcal{C}_{XY}$, then it is an eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ if and only if (3.8) holds. Throughout the following discussion, we always assume that $\lambda \in S_{out}(\varepsilon)$ and $\lambda \geqslant \lambda_+ + \varepsilon$ for a small constant $\varepsilon > 0$. We can write (3.8) as

$$0 = \det\left[\begin{pmatrix} 0 & \mathcal{D}^{-1} \\ \mathcal{D}^{-1} & 0 \end{pmatrix} + \Pi_r(\lambda) + \mathcal{E}_{4r}\right] = \det\left[\begin{pmatrix} \Pi_r^{(1)} & \mathcal{D}^{-1} \\ \mathcal{D}^{-1} & \Pi_r^{(2)} \end{pmatrix} + \mathcal{E}_{4r}\right]. \tag{3.29}$$

Here $\Pi_r(\lambda)$ is a $4r \times 4r$ deterministic matrix defined as

$$\Pi_r(\lambda) := \begin{pmatrix} \Pi_r^{(1)}(\lambda) & 0 \\ 0 & \Pi_r^{(2)}(\lambda) \end{pmatrix}, \tag{3.30}$$

and $\mathcal{E}_{4r}$ is a $4r \times 4r$ random matrix defined as

$$\mathcal{E}_{4r} \equiv \begin{pmatrix} \mathcal{E}_L & \mathcal{E}_{LR} \\ \mathcal{E}_{RL} & \mathcal{E}_R \end{pmatrix} := \begin{pmatrix} \mathbf{U}^\top & 0 \\ 0 & \mathbf{E}^\top \end{pmatrix} (G - \Pi) \begin{pmatrix} \mathbf{U} & 0 \\ 0 & \mathbf{E} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{E}^\top \Pi_R \mathbf{E} - \Pi_r^{(2)} \end{pmatrix}, \tag{3.31}$$

where we have abbreviated (recall the definition of $\mathcal{M}_r$ in (2.22))

$$\Pi_r^{(1)}(\lambda) := \begin{pmatrix} c_1^{-1} m_{1c}(\lambda) I_r & 0 \\ 0 & c_2^{-1} m_{2c}(\lambda) I_r \end{pmatrix}, \quad \Pi_r^{(2)}(\lambda) := \begin{pmatrix} m_{3c}(\lambda) I_r & h(\lambda)\mathcal{M}_r \\ h(\lambda)\mathcal{M}_r^\top & m_{4c}(\lambda) I_r \end{pmatrix},$$

$$\Pi_R(\lambda) := \begin{pmatrix} m_{3c}(\lambda) I_n & h(\lambda) I_n \\ h(\lambda) I_n & m_{4c}(\lambda) I_n \end{pmatrix},$$

and $\mathcal{E}_L$, $\mathcal{E}_{LR}$, $\mathcal{E}_{RL}$ and $\mathcal{E}_R$ are the four $2r \times 2r$ blocks of $\mathcal{E}_{4r}$. Using the large deviation bounds in Theorem B.1 of [14] (cf. Lemma 4.3 below), we can obtain the following approximate isotropic conditions for Z:

$$\|ZZ^\top - I_r\| \prec n^{-1/2}, \quad \text{and} \quad \|Z\mathbf{v}\|_2 \prec n^{-1/2}\|\mathbf{v}\|_2, \tag{3.32}$$

for any deterministic vector $\mathbf{v} \in \mathbb{C}^n$. Using Theorem 3.8 and equation (3.32), we can bound $\mathcal{E}_{4r}$ as

$$\|\mathcal{E}_{4r}\| \prec n^{-1/2}. \tag{3.33}$$

Now using Schur complement formula, we find that (3.29) is equivalent to

$$\det\left[\Pi_r^{(2)} + \mathcal{E}_R - \left(\mathcal{D}^{-1} + \mathcal{E}_{RL}\right) \frac{1}{\Pi_r^{(1)} + \mathcal{E}_L} \left(\mathcal{D}^{-1} + \mathcal{E}_{LR}\right)\right] = 0.$$

Using (3.33) and the first two equations in (3.14), we can reduce this equation to

$$\det\left[\begin{pmatrix} m_{3c}(\lambda)(I_r + \Sigma_a^2) & h(\lambda)\Sigma_a \mathcal{M}_r \Sigma_b \\ h(\lambda)\Sigma_b \mathcal{M}_r^\top \Sigma_a & m_{4c}(\lambda)(I_r + \Sigma_b^2) \end{pmatrix} + \mathcal{E}_{2r} + \mathrm{O}_\prec(n^{-1})\right] = 0. \tag{3.34}$$

Here $\mathcal{E}_{2r}$ is a $2r \times 2r$ random matrix defined as

$$\mathcal{E}_{2r} = \mathcal{D}\mathcal{E}_R\mathcal{D} + (\Pi_r^{(1)})^{-1}\mathcal{E}_L(\Pi_r^{(1)})^{-1} - (\Pi_r^{(1)})^{-1}\mathcal{E}_{LR}\mathcal{D} - \mathcal{D}\mathcal{E}_{RL}(\Pi_r^{(1)})^{-1} = \begin{pmatrix} m_{3c}\mathcal{E}_r^{(1)} & h\mathcal{E}_r^{(2)} \\ h\mathcal{E}_r^{(3)} & m_{4c}\mathcal{E}_r^{(4)} \end{pmatrix},$$

where $\mathcal{E}_r^{(\alpha)}$, $\alpha = 1, 2, 3, 4$, are four $r \times r$ random matrices defined as

$$\begin{aligned} \mathcal{E}_r^{(1)} &= m_{3c}^{-1}\Sigma_a \mathbf{E}_a^\top \left(\mathcal{G}_{(33)} - m_{3c}\right) \mathbf{E}_a \Sigma_a + \Sigma_a \left(\mathbf{E}_a^\top \mathbf{E}_a - I_r\right) \Sigma_a \\ &\quad + m_{3c}\mathbf{U}_a^\top(\mathcal{G}_{(11)} - c_1^{-1}m_{1c})\mathbf{U}_a + \left[\mathbf{U}_a^\top \mathcal{G}_{(13)} \mathbf{E}_a \Sigma_a + \Sigma_a \mathbf{E}_a^\top \mathcal{G}_{(31)} \mathbf{U}_a\right], \\ \mathcal{E}_r^{(2)} &= (\mathcal{E}_r^{(3)})^\top = h^{-1}\Sigma_a \mathbf{E}_a^\top \left(\mathcal{G}_{(34)} - h\right) \mathbf{E}_b \Sigma_b + \Sigma_a \left(\mathbf{E}_a^\top \mathbf{E}_b - \mathcal{M}_r\right) \Sigma_b \\ &\quad + \frac{m_{3c}m_{4c}}{h}\mathbf{U}_a^\top \mathcal{G}_{(12)}\mathbf{U}_b + \frac{m_{3c}}{h}\mathbf{U}_a^\top \mathcal{G}_{(14)} \mathbf{E}_b \Sigma_b + \frac{m_{4c}}{h}\Sigma_a \mathbf{E}_a^\top \mathcal{G}_{(32)}\mathbf{U}_b, \end{aligned}$$

$$\mathcal{E}_r^{(4)} = m_{4c}^{-1}\Sigma_b \mathbf{E}_b^\top \left(\mathcal{G}_{(44)} - m_{4c}\right) \mathbf{E}_b \Sigma_b + \Sigma_b \left(\mathbf{E}_b^\top \mathbf{E}_b - I_r\right) \Sigma_b$$
$$+ m_{4c}\mathbf{U}_b^\top (\mathcal{G}_{(22)} - c_2^{-1} m_{2c})\mathbf{U}_b + \left[\mathbf{U}_b^\top \mathcal{G}_{(24)} \mathbf{E}_b \Sigma_b + \Sigma_b \mathbf{E}_b^\top \mathcal{G}_{(42)} \mathbf{U}_b\right].$$

In these expressions, we have abbreviated the $\mathcal{I}_\alpha \times \mathcal{I}_\beta$ block of $G(z)$ by $\mathcal{G}_{(\alpha\beta)}(z)$ for $\alpha, \beta = 1, 2, 3, 4$, and introduced the notations $\mathbf{E}_a := Z^\top \mathbf{V}_a$ and $\mathbf{E}_b := Z^\top \mathbf{V}_b$. Applying Schur complement formula once again, we obtain that (3.34) is equivalent to

$$\det\left[f_c(\lambda)\left(I_r + \Sigma_a^2\right) + f_c(\lambda)\mathcal{E}_r^{(1)} - \left(\Sigma_a \mathcal{M}_r \Sigma_b + \mathcal{E}_r^{(2)}\right) \frac{1}{I_r + \Sigma_b^2 + \mathcal{E}_r^{(4)}} \left(\Sigma_b \mathcal{M}_r^\top \Sigma_a + \mathcal{E}_r^{(3)}\right) + \mathrm{O}_\prec(n^{-1})\right] = 0,$$

where the function f_c is defined by

$$f_c(z) := \frac{m_{3c}(z) m_{4c}(z)}{h^2(z)} = \frac{z - (c_1 + c_2 - 2c_1c_2) + \sqrt{(z - \lambda_-)(z - \lambda_+)}}{2(1 - c_1)(1 - c_2)}. \tag{3.35}$$

Using $\|\mathcal{E}_r^{(\alpha)}(\lambda)\| \prec n^{-1/2}$, $\alpha = 1, 2, 3, 4$, we can further reduce the above equation to

$$\det\left[f_c(\lambda) I_r - \widehat{\Sigma}_a \mathcal{M}_r \widehat{\Sigma}_b^2 \mathcal{M}_r^\top \widehat{\Sigma}_a + \mathcal{E}_r(\lambda) + \mathrm{O}_\prec(n^{-1})\right] = 0, \tag{3.36}$$

where we recall the notations in (2.22), and $\mathcal{E}_r$ is a $r \times r$ random matrix defined by

$$\mathcal{E}_r := f_c \frac{1}{(I_r + \Sigma_a^2)^{1/2}} \mathcal{E}_r^{(1)} \frac{1}{(I_r + \Sigma_a^2)^{1/2}} + \widehat{\Sigma}_a \mathcal{M}_r \widehat{\Sigma}_b \frac{1}{(I_r + \Sigma_b^2)^{1/2}} \mathcal{E}_r^{(4)} \frac{1}{(I_r + \Sigma_b^2)^{1/2}} \widehat{\Sigma}_b \mathcal{M}_r^\top \widehat{\Sigma}_a$$
$$- \frac{1}{(I_r + \Sigma_a^2)^{1/2}} \mathcal{E}_r^{(2)} \frac{1}{(I_r + \Sigma_b^2)^{1/2}} \widehat{\Sigma}_b \mathcal{M}_r^\top \widehat{\Sigma}_a - \widehat{\Sigma}_a \mathcal{M}_r \widehat{\Sigma}_b \frac{1}{(I_r + \Sigma_b^2)^{1/2}} \mathcal{E}_r^{(3)} \frac{1}{(I_r + \Sigma_a^2)^{1/2}}.$$

Finally, with the SVD (2.23), we can rewrite the equation (3.36) as

$$\det\left[f_c(\lambda) I_r - \mathrm{diag}(t_1, \cdots, t_r) + \mathcal{O}^\top \mathcal{E}_r(\lambda) \mathcal{O} + \mathrm{O}_\prec(n^{-1})\right] = 0. \tag{3.37}$$

It is easy to find the inverse function of f_c in (3.35) when $z \notin [\lambda_-, \lambda_+]$:

$$g_c(\xi) := \xi\left(1 - c_1 + c_1 \xi^{-1}\right)\left(1 - c_2 + c_2 \xi^{-1}\right).$$

Moreover, it is easy to check that $f_c(\lambda_+) = t_c$ for t_c in (1.3). Since $f_c(\lambda)$ is monotonically increasing in λ for $\lambda > \lambda_+$, the function $f_c(\lambda) - t_i = 0$ has a solution in $(\lambda_+, 1)$ if and only if

$$f_c(\lambda_+) < t_i \quad \Leftrightarrow \quad t_c < t_i. \tag{3.38}$$

If (3.38) holds, then t_i gives rise to an outlier lying around $\theta_i = g_c(t_i)$, which explains (2.18). With direct calculations, we can verify the following deterministic estimates on f_c and g_c.

Lemma 3.10 (Lemma 4.1 of [31]). *Fix a large constant $C > 0$. For any $z \in \mathbb{D} := \{z \in \mathbb{C} : \lambda_+ < \mathrm{Re}\, z < C\}$ and $\xi \in f_c(\mathbb{D})$, the following estimates hold:*

$$|f_c(z) - f_c(\lambda_+)| \sim |z - \lambda_+|^{1/2}, \quad |f_c'(z)| \sim |z - \lambda_+|^{-1/2}, \tag{3.39}$$
$$|g_c(\xi) - \lambda_+| \sim |\xi - t_c|^2, \quad |g_c'(\xi)| \sim |\xi - t_c|. \tag{3.40}$$

Now with equation (3.37), we can prove the following proposition, which shows that the limiting law of $n^{1/2}(\widetilde{\lambda}_l - t_l)$ is determined by the limiting law of $n^{1/2}\mathcal{O}^\top \mathcal{E}_r(\theta_l)\mathcal{O}$.

Proposition 3.11 (Reduction to the law of G)**.** *Under the assumptions of Theorem 2.3, fix any $1 \leqslant l \leqslant r$ such that* (2.21) *holds, and define the subset $\gamma(l)$ as in Definition 2.2. Then there exists a constant $\varepsilon > 0$ depending on δ only such that for $1 \leqslant i \leqslant |\gamma(l)|$,*

$$\left|\left(\widetilde{\lambda}_{\alpha(i)} - \theta_l\right) - \mu_i \left\{a(t_l)\left[\operatorname{diag}(t_1, \cdots, t_r) - t_l - \mathcal{O}^\top \mathcal{E}_r(\theta_l)\mathcal{O}\right]_{[\![\gamma(l)]\!]}\right\}\right| \prec n^{-1/2-\varepsilon}, \tag{3.41}$$

where $\alpha : \{1, \cdots, \gamma(l)\} \to \{1, \cdots, r\}$ is a labelling function so that $\widetilde{\lambda}_{\alpha(i)}$ is the i-th largest value in the set $\{\widetilde{\lambda}_i : i \in \gamma(l)\}$, and μ_i is the i-th eigenvalue of the $|\gamma(l)| \times |\gamma(l)|$ matrix

$$a(t_l)\left[\operatorname{diag}(t_1, \cdots, t_r) - t_l - \mathcal{O}^\top \mathcal{E}_r(\theta_l)\mathcal{O}\right]_{[\![\gamma(l)]\!]}$$

in the sense of (2.20)*. Recall that $a(t_l)$ is defined in* (2.28)*.*

Proof. By Lemma 3.3 and the condition (2.21), we have that $\widetilde{\lambda}_l \in S_{out}(\varepsilon)$ and $\widetilde{\lambda}_l \geqslant \lambda_+ + \varepsilon$ with high probability for a sufficiently small constant $\varepsilon > 0$. Thus the above discussion starting at (3.29) will finally lead to the equation (3.37). Armed with (3.1), equation (3.37) and the estimates in Lemma 3.10, the proof is the same as the one for Proposition 4.5 of [27], so we omit the details. In fact, one can easily see why (3.41) holds by performing a Taylor expansion of $f_c(\widetilde{\lambda}_{\alpha(i)})$ around θ_l in (3.37), and noticing that $1/f_c'(\theta_l) = g_c'(t_l) = a(t_l)$. ☐

By Proposition 3.11, to prove Theorem 2.3, it suffices to study the CLT of $\mathcal{O}^\top \mathcal{E}_r(\theta_l)\mathcal{O}$. With a lengthy but straightforward calculation, we can calculate that

$$\mathcal{E}_r(\theta_l) = \mathcal{E}_r^{(1)}(\theta_l) + \mathcal{E}_r^{(2)}(\theta_l), \tag{3.42}$$

where

$$\begin{aligned} \mathcal{E}_r^{(1)}(\theta_l) := f_c(\theta_l)\widehat{\Sigma}_a \mathbf{V}_a^\top \left(ZZ^\top - I_r\right)\mathbf{V}_a\widehat{\Sigma}_a + \widehat{\Sigma}_a\mathcal{M}_r\widehat{\Sigma}_b^2\mathbf{V}_b^\top\left(ZZ^\top - I_r\right)\mathbf{V}_b\widehat{\Sigma}_b^2\mathcal{M}_r^\top\widehat{\Sigma}_a \\ - \widehat{\Sigma}_a\mathbf{V}_a^\top\left(ZZ^\top - I_r\right)\mathbf{V}_b\widehat{\Sigma}_b^2\mathcal{M}_r^\top\widehat{\Sigma}_a - \widehat{\Sigma}_a\mathcal{M}_r\widehat{\Sigma}_b^2\mathbf{V}_b^\top\left(ZZ^\top - I_r\right)\mathbf{V}_a\widehat{\Sigma}_a, \end{aligned} \tag{3.43}$$

and

$$\mathcal{E}_r^{(2)}(\theta_l) := f_c(\theta_l)m_{3c}(\theta_l)\mathfrak{A}^\top(\theta_l)\begin{pmatrix} \mathbf{U}_a^\top & 0 & 0 & 0 \\ 0 & \mathbf{U}_b^\top & 0 & 0 \\ 0 & 0 & Z & 0 \\ 0 & 0 & 0 & Z \end{pmatrix}[G(\theta_l) - \Pi(\theta_l)]\begin{pmatrix} \mathbf{U}_a & 0 & 0 & 0 \\ 0 & \mathbf{U}_b & 0 & 0 \\ 0 & 0 & Z^\top & 0 \\ 0 & 0 & 0 & Z^\top \end{pmatrix}\mathfrak{A}(\theta_l). \tag{3.44}$$

Here $\mathfrak{A}$ is a $4r \times r$ matrix defined by

$$\mathfrak{A}(\theta_l) := \begin{bmatrix} (I_r + \Sigma_a^2)^{-1/2} \\ -h(\theta_l)m_{3c}^{-1}(\theta_l)(1 + \Sigma_b^2)^{-1/2}\widehat{\Sigma}_b\mathcal{M}_r^\top\widehat{\Sigma}_a \\ m_{3c}^{-1}(\theta_l)\mathbf{V}_a\widehat{\Sigma}_a \\ -h(\theta_l)m_{3c}^{-1}(\theta_l)m_{4c}^{-1}(\theta_l)\mathbf{V}_b\widehat{\Sigma}_b^2\mathcal{M}_r^\top\widehat{\Sigma}_a \end{bmatrix}.$$

By classical CLT, we have that

$$\sqrt{n}\left(ZZ^\top - I_r\right) \Rightarrow \Upsilon_z, \tag{3.45}$$

where Υ_z is an $r \times r$ symmetric Gaussian matrix whose entries are independent up to symmetry, and have mean zero and variances (recall the notations in (2.27))

$$\mathbb{E}(\Upsilon_z)_{ij}^2 = 1, \quad i \neq j, \quad \text{and} \quad \mathbb{E}(\Upsilon_z)_{ii}^2 = \mu_z^{(4)} - 1.$$

With this result, we can easily derive the CLT for $\mathcal{E}_r^{(1)}$. Therefore, for the remaining proof, we focus on proving the CLT for the matrix

$$\mathcal{M}_0(\theta_l) := \sqrt{n}\begin{pmatrix} \mathbf{U}_a^\top & 0 & 0 & 0 \\ 0 & \mathbf{U}_b^\top & 0 & 0 \\ 0 & 0 & Z & 0 \\ 0 & 0 & 0 & Z \end{pmatrix}[G(\theta_l) - \Pi(\theta_l)]\begin{pmatrix} \mathbf{U}_a & 0 & 0 & 0 \\ 0 & \mathbf{U}_b & 0 & 0 \\ 0 & 0 & Z^\top & 0 \\ 0 & 0 & 0 & Z^\top \end{pmatrix}. \tag{3.46}$$

As discussed in the introduction, in order to divide the lengthy calculation into several parts that are more manageable, we will adopt the strategy in [26, 27]. First, in Section 4, we prove the CLT for $\mathcal{M}_0(\theta_l)$ in an "almost Gaussian" case, where most of the entries of X and Y are Gaussian. Then, in Section 5, we show that the general case in the setting of Theorem 2.3 is sufficiently close to the almost Gaussian case, thereby completing the proof of Theorem 2.3.

4 The almost Gaussian case

In this section, we calculate the limiting distribution of $\mathcal{M}_0(\theta_l)$ in a special almost Gaussian case. The extension to the general setting in Theorem 2.3 will be postponed to Section 5. We fix a small constant $\tau_0 > 0$ in this section, and use $n^{-\tau_0}$ as a cutoff scale in the entries of $\mathbf{U}_a$ and $\mathbf{U}_b$, below which the corresponding entries of X and Y are Gaussian. Our goal is to prove the following proposition.

Proposition 4.1. *Fix any $1 \leqslant l \leqslant r$ and a sufficiently small constant $\tau_0 > 0$. Suppose Assumption 2.1 and (2.21) hold. Suppose X and Y satisfy that for $k \in \mathcal{I}_1$,*

$$\max_{1\leqslant i\leqslant r} |\mathbf{u}_i^a(k)| \leqslant n^{-\tau_0} \;\Rightarrow\; X_{k\mu} \text{ is Gaussian}, \;\; \mu \in \mathcal{I}_3, \tag{4.1}$$

and for $k \in \mathcal{I}_2$,

$$\max_{1\leqslant i\leqslant r} |\mathbf{u}_i^b(k)| \leqslant n^{-\tau_0} \;\Rightarrow\; Y_{k\mu} \text{ is Gaussian}, \;\; \mu \in \mathcal{I}_4. \tag{4.2}$$

Then for any bounded continuous function f, we have that

$$\lim_n \left[\mathbb{E} f\left(\left(\sqrt{n}\mathcal{O}^\top \mathcal{E}_r(\theta_l)\mathcal{O}\right)_{[\![\gamma(l)]\!]}\right) - \mathbb{E} f(\Upsilon_l)\right] = 0, \tag{4.3}$$

where Υ_l is the Gaussian random matrix defined in Theorem 2.3.

For simplicity, in the proof below we often drop the spectral parameter $z = \theta_l$ from our notations. Using (3.32) and the SVD of Z, we can find an $n \times r$ partial orthogonal matrix V such that

$$V^\top V = I_r, \quad \|V - Z^\top\|_F \prec n^{-1/2}. \tag{4.4}$$

From (3.32) and (4.4), we also obtain the following delocalization estimate:

$$\|V\|_{\max} \leqslant \|Z^\top\|_{\max} + n^{-1/2+\varepsilon/2} \leqslant n^{-1/2+\varepsilon}, \tag{4.5}$$

with high probability for any fixed $\varepsilon > 0$. Now using (4.4) and (3.24), we get that

$$\|\mathcal{M}(\theta_l) - \mathcal{M}_0(\theta_l)\| \prec n^{-1/2}, \tag{4.6}$$

where $\mathcal{M}$ is a $4r \times 4r$ random matrix defined by

$$\mathcal{M}(\theta_l) := \sqrt{n}\begin{pmatrix} \mathbf{U}_a^\top & 0 & 0 & 0 \\ 0 & \mathbf{U}_b^\top & 0 & 0 \\ 0 & 0 & V^\top & 0 \\ 0 & 0 & 0 & V^\top \end{pmatrix}[G(\theta_l) - \Pi(\theta_l)]\begin{pmatrix} \mathbf{U}_a & 0 & 0 & 0 \\ 0 & \mathbf{U}_b & 0 & 0 \\ 0 & 0 & V & 0 \\ 0 & 0 & 0 & V \end{pmatrix}. \tag{4.7}$$

Hence to obtain the CLT of $\mathcal{M}_0(\theta_l)$, it suffices to study $\mathcal{M}(\theta_l)$.

In the proof, we will use the *minors* of H and G defined as follows.

Definition 4.2 (Minors)**.** *For any $\mathcal{J} \times \mathcal{J}$ matrix $\mathcal{A}$ and $\mathbb{T} \subseteq \mathcal{J}$, where $\mathcal{J}$ and $\mathbb{T}$ are some index sets, we define the minor $\mathcal{A}^{(\mathbb{T})} := (\mathcal{A}_{ab} : a, b \in \mathcal{J}\backslash\mathbb{T})$ as the $(\mathcal{J}\backslash\mathbb{T}) \times (\mathcal{J}\backslash\mathbb{T})$ matrix obtained by removing all rows and columns indexed by $\mathbb{T}$. Note that we keep the names of indices when defining $\mathcal{A}^{(\mathbb{T})}$, i.e. $(\mathcal{A}^{(\mathbb{T})})_{ab} = \mathcal{A}_{ab}$ for $a, b \notin \mathbb{T}$. Correspondingly, we define the resolvent minor as $G^{(\mathbb{T})}(z) := [H^{(\mathbb{T})}(z)]^{-1}$. For convenience, we will adopt the convention that $\mathcal{A}^{(T)}_{\mathfrak{a}\mathfrak{b}} = 0$ if $\mathfrak{a} \in \mathbb{T}$ or $\mathfrak{b} \in \mathbb{T}$. We will abbreviate $(\{\mathfrak{a}\}) \equiv (\mathfrak{a})$, $(\{\mathfrak{a}, \mathfrak{b}\}) \equiv (\mathfrak{a}\mathfrak{b})$ and $\sum_{\mathfrak{a}}^{(\mathbb{T})} := \sum_{\mathfrak{a}\notin\mathbb{T}}$.*

Recall the following large deviation bounds proved in [14] for linear and quadratic forms of independent random variables satisfying (2.8).

Lemma 4.3 (Theorem B.1 of [14])**.** *Let (x_i), (y_j) be independent families of centered independent random variables, and $(\mathcal{A}_i)$, $(\mathcal{B}_{ij})$ be families of deterministic complex numbers. Suppose the entries x_i, y_j have variances at most n^{-1} and satisfy (2.8). Then the following large deviation bounds hold:*

$$\Big|\sum_i \mathcal{A}_i x_i\Big| \prec \frac{1}{\sqrt{n}}\Big(\sum_i |\mathcal{A}_i|^2\Big)^{1/2}, \quad \Big|\sum_{i,j} x_i \mathcal{B}_{ij} y_j\Big| \prec \frac{1}{n}\Big(\sum_{i,j} |\mathcal{B}_{ij}|^2\Big)^{1/2}, \quad \Big|\sum_{i\neq j} x_i \mathcal{B}_{ij} x_j\Big| \prec \frac{1}{n}\Big(\sum_{i\neq j} |\mathcal{B}_{ij}|^2\Big)^{1/2}.$$

For convenience, we introduce the following shorthand for the equivalence relation between two random vectors of fixed size in the sense of asymptotic distributions.

Definition 4.4. *For two sequences of random vectors $\mathcal{A}_n$ and $\mathcal{B}_n$ in $\mathbb{R}^k$, where $k \in \mathbb{N}$ is a fixed integer, we write $\mathcal{A}_n \overset{d}{\sim} \mathcal{B}_n$ if*

$$\lim_{n\to\infty} [\mathbb{E}f(\mathcal{A}_n) - \mathbb{E}f(\mathcal{B}_n)] = 0$$

for any bounded continuous function f.

In the proof, we will frequently use the following simple fact, which can be proved easily using characteristic functions. Given two sequences of random vectors $\mathcal{A}_n$ and $\mathcal{B}_n$, suppose that conditioning on $\mathcal{A}_n$, we have $\mathcal{B}_n \overset{d}{\sim} \mathcal{D}_n$, where $\mathcal{D}_n$ has an asymptotic distribution that does not depend on $\mathcal{A}_n$. Then we have that

$$\mathcal{A}_n + \mathcal{B}_n \overset{d}{\sim} \mathcal{A}_n + \mathcal{D}_n, \tag{4.8}$$

where on the right-hand side $\mathcal{D}_n$ is taken to be independent of $\mathcal{A}_n$. One immediate use of this fact is to decouple the randomness of $\mathcal{M}(\theta_l)$ from that of Z as long as we can show that the limiting distribution of $\mathcal{M}(\theta_l)$ does not depend on Z. In the following proof, we will condition on Z and V, i.e. they are regarded as deterministic matrices unless specified otherwise.

4.1 Step 1: Rewriting $\mathcal{M}(x)$

We start with some linear algebra to write $\mathcal{M}(x)$ into a form that is more amenable to our analysis. Our main tool is the rotational invariance of multivariate Gaussian distributions. Note that since $\|\mathbf{u}_i^a\|_2 = 1$ and $\|\mathbf{u}_i^b\|_2 = 1$ for $1 \leqslant i \leqslant r$, we have

$$\Big|\Big\{k : \max_{1\leqslant i\leqslant r} |u_i^a(k)| > n^{-\tau_0}\Big\}\Big| \leqslant rn^{2\tau_0}, \quad \Big|\Big\{k : \max_{1\leqslant i\leqslant r} |u_i^b(k)| > n^{-\tau_0}\Big\}\Big| \leqslant rn^{2\tau_0}.$$

We first permute the rows of $\mathbf{U}_a$, $\mathbf{U}_b$, X and Y according to

$$\mathcal{M}(\theta_l) = \sqrt{n}\begin{pmatrix} \mathbf{U}_a^\top O_1^\top & 0 & 0 & 0 \\ 0 & \mathbf{U}_b^\top O_2^\top & 0 & 0 \\ 0 & 0 & V^\top & 0 \\ 0 & 0 & 0 & V^\top \end{pmatrix}\begin{pmatrix} O_1 & 0 & 0 & 0 \\ 0 & O_2 & 0 & 0 \\ 0 & 0 & I_n & 0 \\ 0 & 0 & 0 & I_n \end{pmatrix}$$
$$\times [G(\theta_l) - \Pi(\theta_l)] \begin{pmatrix} O_1^\top & 0 & 0 & 0 \\ 0 & O_2^\top & 0 & 0 \\ 0 & 0 & I_n & 0 \\ 0 & 0 & 0 & I_n \end{pmatrix}\begin{pmatrix} O_1\mathbf{U}_a & 0 & 0 & 0 \\ 0 & O_2\mathbf{U}_b & 0 & 0 \\ 0 & 0 & V & 0 \\ 0 & 0 & 0 & V \end{pmatrix},$$

where O_1 and O_2 are $p \times p$ and $q \times q$ permutation matrices. After permuting the rows of $\mathbf{U}_a$ and $\mathbf{U}_b$ and renaming the matrices if necessary, we can assume that $\mathbf{U}_a$ and $\mathbf{U}_b$ take the forms

$$\mathbf{U}_a = \begin{pmatrix} \mathbf{O}_1 \\ \mathbf{W}_1 \end{pmatrix}, \quad \mathbf{U}_b = \begin{pmatrix} \mathbf{O}_2 \\ \mathbf{W}_2 \end{pmatrix}, \tag{4.9}$$

where for some integer $\rho \leqslant rn^{2\tau_0}$, the following properties hold:

(i) $\mathbf{O}_1$, $\mathbf{O}_2$ are $\rho \times r$ matrices, $\mathbf{W}_1$ is a $(p-\rho) \times r$ matrix, and $\mathbf{W}_2$ is a $(q-\rho) \times r$ matrix,

(ii) $\|\mathbf{W}_1\|_{\max} \leqslant n^{-\tau_0}$ and $\|\mathbf{W}_2\|_{\max} \leqslant n^{-\tau_0}$.

On the other hand, we have

$$\begin{pmatrix} O_1 & 0 & 0 & 0 \\ 0 & O_2 & 0 & 0 \\ 0 & 0 & I_n & 0 \\ 0 & 0 & 0 & I_n \end{pmatrix}[G(\theta_l) - \Pi(\theta_l)]\begin{pmatrix} O_1^\top & 0 & 0 & 0 \\ 0 & O_2^\top & 0 & 0 \\ 0 & 0 & I_n & 0 \\ 0 & 0 & 0 & I_n \end{pmatrix}$$
$$= \begin{bmatrix} 0 & \begin{pmatrix} O_1 X & 0 \\ 0 & O_2 Y \end{pmatrix} \\ \begin{pmatrix} X^\top O_1^\top & 0 \\ 0 & Y^\top O_2^\top \end{pmatrix} & \begin{pmatrix} \theta_l I_n & \theta_l^{1/2} I_n \\ \theta_l^{1/2} I_n & \theta_l I_n \end{pmatrix}^{-1} \end{bmatrix}^{-1} - \Pi(\theta_l),$$

We rename the permuted matrices O_1X and O_2Y as X and Y. Then X and Y take the forms

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},$$

where X_1 and Y_1 are $\rho \times n$ matrices, X_2 is a $(p-\rho) \times n$ Gaussian matrix and Y_2 is a $(q-\rho) \times n$ Gaussian matrix. Next we rotate $\mathbf{W}_1$ and $\mathbf{W}_2$ using orthogonal $(p-\rho) \times (p-\rho)$ and $(q-\rho) \times (q-\rho)$ matrices $\widetilde{S}_1$ and $\widetilde{S}_2$ so that

$$\widetilde{S}_1^\top \mathbf{W}_1 = \begin{pmatrix} \widetilde{\mathbf{W}}_1 \\ 0 \end{pmatrix}, \quad \widetilde{S}_2^\top \mathbf{W}_2 = \begin{pmatrix} \widetilde{\mathbf{W}}_2 \\ 0 \end{pmatrix},$$

where $\widetilde{\mathbf{W}}_1$ and $\widetilde{\mathbf{W}}_2$ are $r \times r$ matrices satisfying that

$$\mathbf{O}_\alpha^\top \mathbf{O}_\alpha + \mathbf{W}_\alpha^\top \mathbf{W}_\alpha = \mathbf{O}_\alpha^\top \mathbf{O}_\alpha + \widetilde{\mathbf{W}}_\alpha^\top \widetilde{\mathbf{W}}_\alpha = I_r, \quad \alpha = 1, 2. \tag{4.10}$$

Similarly, we rotate V using an orthogonal $n \times n$ matrix $\widetilde{S} = \big(V, S\big)$, where S is an $n \times (n-r)$ matrix satisfying $S^\top S = I_{n-r}$ and $S^\top V = 0$.

With the above notations, we can rewrite $\mathcal{M}$ in (4.7) as

$$\mathcal{M}=\sqrt{n}\begin{pmatrix} \widetilde{\mathbf{U}}_a^\top & 0 & 0 & 0 \\ 0 & \widetilde{\mathbf{U}}_b^\top & 0 & 0 \\ 0 & 0 & \widetilde{\mathbf{V}}^\top & 0 \\ 0 & 0 & 0 & \widetilde{\mathbf{V}}^\top \end{pmatrix} \begin{bmatrix} 0 & \begin{pmatrix} \widetilde{X} & 0 \\ 0 & \widetilde{Y} \end{pmatrix} \\ \begin{pmatrix} \widetilde{X}^\top & 0 \\ 0 & \widetilde{Y}^\top \end{pmatrix} & \begin{pmatrix} \theta_l I_n & \theta_l^{1/2} I_n \\ \theta_l^{1/2} I_n & \theta_l I_n \end{pmatrix}^{-1} \end{bmatrix}^{-1} \begin{pmatrix} \widetilde{\mathbf{U}}_a & 0 & 0 & 0 \\ 0 & \widetilde{\mathbf{U}}_b & 0 & 0 \\ 0 & 0 & \widetilde{\mathbf{V}} & 0 \\ 0 & 0 & 0 & \widetilde{\mathbf{V}} \end{pmatrix},$$

where we have abbreviated that

$$\widetilde{\mathbf{U}}_a := \begin{pmatrix} \mathbf{O}_1 \\ \widetilde{\mathbf{W}}_1 \\ 0 \end{pmatrix}, \quad \widetilde{\mathbf{U}}_b := \begin{pmatrix} \mathbf{O}_2 \\ \widetilde{\mathbf{W}}_2 \\ 0 \end{pmatrix}, \quad \widetilde{\mathbf{V}} := \begin{pmatrix} I_r \\ 0 \end{pmatrix}, \quad \widetilde{X} := \begin{pmatrix} I_\rho & 0 \\ 0 & \widetilde{S}_1^\top \end{pmatrix} X\widetilde{S}, \quad \widetilde{Y} := \begin{pmatrix} I_\rho & 0 \\ 0 & \widetilde{S}_2^\top \end{pmatrix} Y\widetilde{S}.$$

Using the rotational invariance of X_2, we can write $\widetilde{X}$ as

$$\widetilde{X} \overset{d}{=} \begin{pmatrix} X_1 V, X_1 S \\ X_2 \end{pmatrix} \equiv \begin{pmatrix} X_1 V & X_1 S \\ X_{RL}^{(1)} & X_R^{(1)} \\ X_{RL}^{(2)} & X_R^{(2)} \end{pmatrix},$$

where "$\overset{d}{=}$" means "equal in distribution", and $X_{RL}^{(1)}$, $X_{RL}^{(2)}$, $X_R^{(1)}$ and $X_R^{(2)}$ are respectively $r\times r$, $(p-\rho-r)\times r$, $r\times(n-r)$ and $(p-\rho-r)\times(n-r)$ Gaussian matrices. We have a similar decomposition for Y:

$$\widetilde{Y} \overset{d}{=} \begin{pmatrix} Y_1 V, Y_1 S \\ Y_2 \end{pmatrix} \equiv \begin{pmatrix} Y_1 V & Y_1 S \\ Y_{RL}^{(1)} & Y_R^{(1)} \\ Y_{RL}^{(2)} & Y_R^{(2)} \end{pmatrix}.$$

For simplicity of notations, we denote $\widetilde{r} = r + \rho$ and

$$\mathbb{T} := \{1, \cdots, \widetilde{r}\} \cup \{p+1, \cdots, p+\widetilde{r}\} \cup \{p+q+1, \cdots, p+q+r\} \cup \{p+q+n+1, \cdots, p+q+n+r\}.$$

Then using Schur's complement formula, we obtain that

$$\mathcal{M} \overset{d}{=} \sqrt{n}\,\mathbf{O}^\top \mathcal{B}^{-1} \mathbf{O}. \tag{4.11}$$

Here $\mathbf{O}$ is a $(2\widetilde{r}+2r)\times 4r$ matrix defined as

$$\mathbf{O} := \begin{bmatrix} \begin{pmatrix} \mathbf{O}_1 \\ \widetilde{\mathbf{W}}_1 \end{pmatrix} & 0 & 0 & 0 \\ 0 & \begin{pmatrix} \mathbf{O}_2 \\ \widetilde{\mathbf{W}}_2 \end{pmatrix} & 0 & 0 \\ 0 & 0 & I_r & 0 \\ 0 & 0 & 0 & I_r \end{bmatrix},$$

and $\mathcal{B}$ is a $(2\widetilde{r}+2r)\times(2\widetilde{r}+2r)$ symmetric matrix defined as (recall Definition 4.2)

$$\mathcal{B} := \begin{bmatrix} 0 \cdot I_{2\widetilde{r}} & 0 \\ 0 & \begin{pmatrix} \theta_l I_r & \theta_l^{1/2} I_r \\ \theta_l^{1/2} I_r & \theta_l I_r \end{pmatrix}^{-1} \end{bmatrix} + H_1 - F^\top G^{(\mathbb{T})}(\theta_l) F,$$

where H_1 and F are defined by

$$H_1 := \begin{bmatrix} 0 & \begin{pmatrix} \begin{pmatrix} X_1 V \\ X_{RL}^{(1)} \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} Y_1 V \\ Y_{RL}^{(1)} \end{pmatrix} \end{pmatrix} \\ \begin{pmatrix} \left(V^\top X_1^\top, (X_{RL}^{(1)})^\top\right) & 0 \\ 0 & \left(V^\top Y_1^\top, (Y_{RL}^{(1)})^\top\right) \end{pmatrix} & 0 \end{bmatrix},$$

and

$$F := \begin{bmatrix} 0 & \begin{pmatrix} X_{RL}^{(2)} & 0 \\ 0 & Y_{RL}^{(2)} \end{pmatrix} \\ \begin{pmatrix} \left(S^\top X_1^\top, (X_R^{(1)})^\top\right) & 0 \\ 0 & \left(S^\top Y_1^\top, (Y_R^{(1)})^\top\right) \end{pmatrix} & 0 \end{bmatrix}.$$

Using (3.19), we can rewrite $\mathcal{B}$ as

$$\mathcal{B} := \widetilde{\Pi}_{\widetilde{r},r}^{-1} + H_1 + \begin{pmatrix} m_{3c} I_{\widetilde{r}} & 0 & 0 & 0 \\ 0 & m_{4c} I_{\widetilde{r}} & 0 & 0 \\ 0 & 0 & m_{1c} I_r & 0 \\ 0 & 0 & 0 & m_{2c} I_r \end{pmatrix} - F^\top \Pi^{(\mathbb{T})} F - F^\top (G^{(\mathbb{T})} - \Pi^{(\mathbb{T})}) F, \tag{4.12}$$

where $\Pi^{(\mathbb{T})}$ is the minor of Π as defined in Definition 4.2 and $\widetilde{\Pi}_{\widetilde{r},r}$ is defined by

$$\widetilde{\Pi}_{\widetilde{r},r} := \begin{bmatrix} \begin{pmatrix} c_1^{-1} m_{1c} I_{\widetilde{r}} & 0 \\ 0 & c_2^{-1} m_{2c} I_{\widetilde{r}} \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} m_{3c} I_r & h I_r \\ h I_r & m_{4c} I_r \end{pmatrix} \end{bmatrix}. \tag{4.13}$$

4.2 Step 2: Concentration estimates

In this step, we establish some (almost) sharp concentration estimates on the terms on the right-hand side of (4.12). More precisely, we claim that

$$F^\top F - \begin{pmatrix} I_{\widetilde{r}} & 0 & 0 & 0 \\ 0 & I_{\widetilde{r}} & 0 & 0 \\ 0 & 0 & c_1 I_r & 0 \\ 0 & 0 & 0 & c_2 I_r \end{pmatrix} = \mathrm{O}_\prec(n^{-1/2+2\tau_0}), \tag{4.14}$$

and

$$F^\top \Pi^{(\mathbb{T})} F - \mathbb{E}_F(F^\top \Pi^{(\mathbb{T})} F) = \mathrm{O}_\prec(n^{-1/2+2\tau_0}), \tag{4.15}$$

where $\mathbb{E}_F$ denotes the partial expectation over the randomness in F. (To avoid confusion, we emphasize that the matrix S is regarded as a deterministic matrix because of the conditioning on Z. Hence the randomness in F does not include the randomness in S.) Using the facts $S^\top S = I_{n-r}$ and $\widetilde{r} = \mathrm{O}(n^{2\tau_0})$, we can get that

$$\mathbb{E}_F\left(F^\top \Pi^{(\mathbb{T})} F\right) = \begin{pmatrix} m_{3c} I_{\widetilde{r}} & 0 & 0 & 0 \\ 0 & m_{4c} I_{\widetilde{r}} & 0 & 0 \\ 0 & 0 & m_{1c} I_r & 0 \\ 0 & 0 & 0 & m_{2c} I_r \end{pmatrix} + \mathrm{O}\left(n^{-1+2\tau_0}\right). \tag{4.16}$$

The two estimates (4.14) and (4.15) both follow from Lemma 4.3. We consider the terms $(X_{RL}^{(2)})^\top X_{RL}^{(2)}$, $X_1 SS^\top X_1^\top$ and $X_1 S(X_R^{(1)})^\top$ as examples. For $p+q+1 \leqslant \mu, \nu \leqslant p+q+r$, we have that

$$\left|\left[(X_{RL}^{(2)})^\top X_{RL}^{(2)}\right]_{\mu\nu} - \frac{p-\widetilde{r}}{n}\delta_{\mu\nu}\right| = \Big|\sum_{\widetilde{r}+1\leqslant i\leqslant p}\left(X_{i\mu}X_{i\nu} - n^{-1}\delta_{\mu\nu}\right)\Big| \prec \mathrm{O}(n^{-1/2}).$$

For $1 \leqslant i \leqslant \rho$, we have that

$$\begin{aligned}\left|\left(X_1 SS^\top X_1^\top\right)_{ii} - n^{-1}\operatorname{Tr}\left(SS^\top\right)\right| &= \Big|\sum_{\mu\neq\nu\in\mathcal{I}_3} X_{i\mu}X_{i\nu}(SS^\top)_{\mu\nu}\Big| + \Big|\sum_{\mu\in\mathcal{I}_3}(X_{i\mu}^2 - n^{-1})(SS^\top)_{\mu\mu}\Big| \\ &\prec \frac{1}{n}\Big(\sum_{\mu\neq\nu\in\mathcal{I}_3}[(SS^\top)_{\mu\nu}]^2\Big)^{1/2} + \frac{1}{n}\Big(\sum_{\mu\in\mathcal{I}_3}[(SS^\top)_{\mu\mu}]^2\Big)^{1/2} \leqslant \frac{2}{n}\left\{\operatorname{Tr}\left[(SS^\top)^2\right]\right\}^{1/2} = \mathrm{O}(n^{-1/2}),\end{aligned}$$

while for $1 \leqslant i < j \leqslant \rho$, we have that

$$(X_1 SS^\top X_1^\top)_{ij} = \sum_{\mu,\nu\in\mathcal{I}_3} X_{i\mu}X_{j\nu}(SS^\top)_{\mu\nu} \prec \frac{1}{n}\Big(\sum_{\mu,\nu\in\mathcal{I}_3}[(SS^\top)_{\mu\nu}]^2\Big)^{1/2} = \frac{1}{n}\left\{\operatorname{Tr}\left[(SS^\top)^2\right]\right\}^{1/2} = \mathrm{O}(n^{-1/2}).$$

Using the fact that $\operatorname{Tr}(SS^\top) = n-r$, the above two estimates actually give the estimate

$$\left|(X_1 SS^\top X_1^\top)_{ij} - \delta_{ij}\right| \prec n^{-1/2}, \quad 1 \leqslant i,j \leqslant \rho.$$

Finally, for $1 \leqslant i \leqslant \rho$ and $\rho+1 \leqslant j \leqslant \rho + r$, we have that

$$\left[X_1 S(X_R^{(1)})^\top\right]_{ij} = \sum_{\mu,\nu\in\mathcal{I}_3} X_{i\mu}X_{j\nu}S_{\mu\nu} \prec \frac{1}{n}\Big(\sum_{\mu,\nu\in\mathcal{I}_3} S_{\mu\nu}^2\Big)^{1/2} = \frac{1}{n}\left[\operatorname{Tr}(SS^\top)\right]^{1/2} = \mathrm{O}(n^{-1/2}).$$

With similar arguments as above, using Lemma 4.3 we can obtain the following large deviation estimates: for any constant $\varepsilon > 0$,

$$\begin{aligned}&\left\|(X_{RL}^{(2)})^\top X_{RL}^{(2)} - c_1 I_r\right\|_{\max} \leqslant n^{-1/2+\varepsilon}, \quad \left\|(Y_{RL}^{(2)})^\top Y_{RL}^{(2)} - c_2 I_r\right\|_{\max} \leqslant n^{-1/2+\varepsilon}, \\ &\left\|X_1 SS^\top X_1^\top - I_\rho\right\|_{\max} \leqslant n^{-1/2+\varepsilon}, \quad \left\|Y_1 SS^\top Y_1^\top - I_\rho\right\|_{\max} \leqslant n^{-1/2+\varepsilon}, \quad \left\|X_1 SS^\top Y_1\right\|_{\max} \leqslant n^{-1/2+\varepsilon}, \\ &\left\|X_R^{(1)}(X_R^{(1)})^\top - I_r\right\|_{\max} \leqslant n^{-1/2+\varepsilon}, \quad \left\|Y_R^{(1)}(Y_R^{(1)})^\top - I_r\right\|_{\max} \leqslant n^{-1/2+\varepsilon}, \\ &\left\|X_R^{(1)}(Y_R^{(1)})^\top\right\|_{\max} \leqslant n^{-1/2+\varepsilon}, \quad \left\|X_1 S(X_R^{(1)})^\top\right\|_{\max} \leqslant n^{-1/2+\varepsilon}, \quad \left\|X_1 S(Y_R^{(1)})^\top\right\|_{\max} \leqslant n^{-1/2+\varepsilon}, \\ &\left\|Y_1 S(X_R^{(1)})^\top\right\|_{\max} \leqslant n^{-1/2+\varepsilon}, \quad \left\|Y_1 S(Y_R^{(1)})^\top\right\|_{\max} \leqslant n^{-1/2+\varepsilon},\end{aligned} \tag{4.17}$$

with high probability. These estimates immediately imply (4.14) and (4.15) by bounding the operator norms of the error matrices by their Frobenius norms.

By (4.14), we have that $\|F\| = \mathrm{O}(1)$ with high probability. Then using the local law (3.24) and the fact that F is independent of $G^{(\mathbb{T})}$, we get that

$$\left\|F^\top(G^{(\mathbb{T})} - \Pi^{(\mathbb{T})})F\right\| \leqslant (2\widetilde{r} + 2r)\left\|F^\top(G^{(\mathbb{T})} - \Pi^{(\mathbb{T})})F\right\|_{\max} \prec n^{-1/2+2\tau_0}.$$

Under the moment assumption (2.8), each entry in H_1 is of order $\mathrm{O}_\prec(n^{-1/2})$ by Markov's inequality, so we also have that

$$\|H_1\| \leqslant (2\widetilde{r} + 2r)\|H_1\|_{\max} \prec n^{-1/2+2\tau_0}.$$

Finally, by (4.15) and (4.16) we have that

$$\left\| \begin{pmatrix} m_{3c} I_{\widetilde{r}} & 0 & 0 & 0 \\ 0 & m_{4c} I_{\widetilde{r}} & 0 & 0 \\ 0 & 0 & m_{1c} I_r & 0 \\ 0 & 0 & 0 & m_{2c} I_r \end{pmatrix} - F^\top \Pi^{(\mathbb{T})} F \right\| \prec n^{-1/2+2\tau_0}.$$

Hence for $\mathcal{M}$ in (4.11), taking the inverse of (4.12) and performing a simple Taylor expansion we get that

$$\mathcal{M} \stackrel{d}{=} \sqrt{n}\, \mathbf{O}^\top \widetilde{\Pi}_{\widetilde{r},r} \left[-H_1 + (1-\mathbb{E}_F) \left(F^\top \Pi^{(\mathbb{T})} F \right) + F^\top (G^{(\mathbb{T})} - \Pi^{(\mathbb{T})}) F \right] \widetilde{\Pi}_{\widetilde{r},r}\, \mathbf{O} + \mathrm{O}_\prec(n^{-1/2+4\tau_0}), \qquad (4.18)$$

where we also used (4.16) in the derivation. Since τ_0 can be taken as small as possible, it suffices to study the CLT of the first term in (4.18).

4.3 Step 3: CLT of the resolvent

In this step, we establish the CLT of the resolvent term $F^\top (G^{(\mathbb{T})} - \Pi^{(\mathbb{T})}) F$ in (4.18). Conditioning on F, We have the following lemma, whose proof is postponed to Appendix A.

Lemma 4.5. *Fix any F such that the estimates in* (4.17) *holds for a small enough constant $\varepsilon > 0$. Then we have that (recall Definition 4.4)*

$$\sqrt{n}\, \mathbf{O}^\top F^\top (G^{(\mathbb{T})} - \Pi^{(\mathbb{T})}) F\, \mathbf{O} \stackrel{d}{\sim} \begin{pmatrix} a_{11} g_{11} & a_{12} g_{12} & a_{13} g_{13} & a_{14} g_{14} \\ a_{21} g_{21} & a_{22} g_{22} & a_{23} g_{23} & a_{24} g_{24} \\ a_{31} g_{31} & a_{32} g_{32} & a_{33} g_{33} & a_{34} g_{34} \\ a_{41} g_{41} & a_{42} g_{42} & a_{43} g_{43} & a_{44} g_{44} \end{pmatrix}. \qquad (4.19)$$

Here $g_{\alpha\beta}$, $1 \leqslant \alpha \leqslant \beta \leqslant 4$, are independent Gaussian matrices satisfying the following properties: $g_{\alpha\beta} = g_{\beta\alpha}^\top$, $1 \leqslant \alpha < \beta \leqslant 4$, are $r \times r$ random matrices with i.i.d. Gaussian entries $(g_{\alpha\beta})_{ij} \sim \mathcal{N}(0,1)$; $g_{\alpha\alpha}$, $1 \leqslant \alpha \leqslant 4$, are $r \times r$ symmetric GOE (Gaussian orthogonal ensemble) with entries $(g_{\alpha\alpha})_{ij} \sim \mathcal{N}(0, 1+\delta_{ij})$. Moreover, the coefficients are given by

$$\begin{aligned} & a_{11} := m_{3c} \sqrt{\frac{a_c^2 + c_1}{1-c_1} + \frac{a_c^2}{c_1}}, \quad a_{12} = a_{21} := h \sqrt{\frac{a_c^2}{c_2} t_l^2 + \frac{a_c^2 + c_2}{1-c_2}}, \quad a_{13} = a_{31} := \sqrt{\frac{a_c^2 + c_1}{1-c_1}}, \\ & a_{14} = a_{41} := \frac{a_c}{\sqrt{c_1}} \frac{m_{3c}}{h}, \quad a_{22} := m_{4c} \sqrt{\frac{a_c^2 + c_2}{1-c_2} + \frac{a_c^2}{c_2}}, \quad a_{23} = a_{32} := \frac{a_c}{\sqrt{c_2}} \frac{m_{4c}}{h}, \\ & a_{24} = a_{42} := \sqrt{\frac{a_c^2 + c_2}{1-c_2}}, \quad a_{33} := m_{3c}^{-1} \sqrt{c_1 \frac{a_c^2 + c_1}{1-c_1}}, \quad a_{34} = a_{43} := \frac{a_c}{h}, \quad a_{44} := m_{4c}^{-1} \sqrt{c_2 \frac{a_c^2 + c_2}{1-c_2}}, \end{aligned} \qquad (4.20)$$

where we have introduced the notation

$$a_c^2 := \frac{t_c^2}{t_l^2 - t_c^2}. \qquad (4.21)$$

With Lemma 4.5, we get the weak convergence

$$\sqrt{n}\, \mathbf{O}^\top \widetilde{\Pi}_{\widetilde{r},r} F^\top (G^{(\mathbb{T})} - \Pi^{(\mathbb{T})}) F \widetilde{\Pi}_{\widetilde{r},r}\, \mathbf{O} \quad \Rightarrow \quad \widetilde{\Pi}_{r,r} \begin{pmatrix} a_{11} g_{11} & a_{12} g_{12} & a_{13} g_{13} & a_{14} g_{14} \\ a_{21} g_{21} & a_{22} g_{22} & a_{23} g_{23} & a_{24} g_{24} \\ a_{31} g_{31} & a_{32} g_{32} & a_{33} g_{33} & a_{34} g_{34} \\ a_{41} g_{41} & a_{42} g_{42} & a_{43} g_{43} & a_{44} g_{44} \end{pmatrix} \widetilde{\Pi}_{r,r}, \qquad (4.22)$$

using the simple identity $\widetilde{\Pi}_{\tilde{r},r}\,\mathbf{O} = \mathbf{O}\,\widetilde{\Pi}_{r,r}$, with $\widetilde{\Pi}_{r,r}$ being a $4r\times 4r$ matrix defined in a similar way as (4.13):

$$\widetilde{\Pi}_{r,r} := \begin{bmatrix} \begin{pmatrix} c_1^{-1}m_{1c}I_r & 0 \\ 0 & c_2^{-1}m_{2c}I_r \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} m_{3c}I_r & hI_r \\ hI_r & m_{4c}I_r \end{pmatrix} \end{bmatrix}. \tag{4.23}$$

4.4 Step 4: Calculating the limiting covariances

In this step, we expand (4.18) and show a CLT for each term. The main work is to calculate the limiting covariances. First, Lemma 4.5 already gives the CLT for $\sqrt{n}\,\mathbf{O}^\top\,\widetilde{\Pi}_{\tilde{r},r}F^\top(G^{(\mathbb{T})}-\Pi^{(\mathbb{T})})F\widetilde{\Pi}_{\tilde{r},r}\,\mathbf{O}$. We still need to study the term

$$\sqrt{n}\,\mathbf{O}^\top\,\widetilde{\Pi}_{\tilde{r},r}\left[-H_1+(1-\mathbb{E}_F)\left(F^\top\Pi^{(\mathbb{T})}F\right)\right]\widetilde{\Pi}_{\tilde{r},r}\,\mathbf{O} = \widetilde{\Pi}_{r,r}Q_{4r}\widetilde{\Pi}_{r,r} = \widetilde{\Pi}_{r,r}\begin{pmatrix} Q_L & Q_{LR} \\ Q_{RL} & Q_R \end{pmatrix}\widetilde{\Pi}_{r,r},$$

where Q_{4r} is a $4r\times 4r$ symmetric matrix, with Q_L, Q_{LR}, Q_{RL} and Q_R being the $2r\times 2r$ blocks defined by

$$\begin{aligned} Q_L &:= \begin{pmatrix} Q_L^{(1)} & Q_L^{(2)} \\ Q_L^{(3)} & Q_L^{(4)} \end{pmatrix}, \quad Q_R := \sqrt{n}\begin{pmatrix} -m_{3c}^{-1}\mathbb{IE}(X_{RL}^{(2)})^\top X_{RL}^{(2)} & 0 \\ 0 & -m_{4c}^{-1}\mathbb{IE}(Y_{RL}^{(2)})^\top Y_{RL}^{(2)} \end{pmatrix}, \\ Q_{LR} = Q_{RL}^\top &:= \sqrt{n}\begin{pmatrix} -\,\mathbf{O}_1^\top\,X_1V-\widetilde{\mathbf{W}}_1^\top X_{RL}^{(1)} & 0 \\ 0 & -\,\mathbf{O}_2^\top\,Y_1V-\widetilde{\mathbf{W}}_2^\top Y_{RL}^{(1)} \end{pmatrix}. \end{aligned}$$

Here we have abbreviated $\mathbb{IE} := 1-\mathbb{E}_F$, and the four $r\times r$ blocks of Q_L are defined as

$$\begin{aligned} Q_L^{(1)} &:= \sqrt{n}\mathbb{IE}\left[m_{3c}\left(\mathbf{O}_1^\top\,X_1S+\widetilde{\mathbf{W}}_1^\top X_R^{(1)}\right)\left(S^\top X_1^\top\,\mathbf{O}_1+(X_R^{(1)})^\top\widetilde{\mathbf{W}}_1\right)\right], \\ Q_L^{(2)} = (Q_L^{(3)})^\top &:= \sqrt{n}\mathbb{IE}\left[h\left(\mathbf{O}_1^\top\,X_1S+\widetilde{\mathbf{W}}_1^\top X_R^{(1)}\right)\left(S^\top Y_1^\top\,\mathbf{O}_2+(Y_R^{(1)})^\top\widetilde{\mathbf{W}}_2\right)\right], \\ Q_L^{(4)} &:= \sqrt{n}\mathbb{IE}\left[m_{4c}\left(\mathbf{O}_2^\top\,Y_1S+\widetilde{\mathbf{W}}_2^\top Y_R^{(1)}\right)\left(S^\top Y_1^\top\,\mathbf{O}_2+(Y_R^{(1)})^\top\widetilde{\mathbf{W}}_2\right)\right]. \end{aligned}$$

Now using (3.44), (4.6), (4.18), (4.22) and the simple fact (4.8), we obtain that

$$\sqrt{n}\mathcal{O}^\top\mathcal{E}_r^{(2)}\mathcal{O} \overset{d}{\sim} t_l m_{3c}\,\mathbf{E}_{AB}^\top\begin{pmatrix} a_{11}g_{11} & a_{12}g_{12} & a_{13}g_{13} & a_{14}g_{14} \\ a_{21}g_{21} & a_{22}g_{22} & a_{23}g_{23} & a_{24}g_{24} \\ a_{31}g_{31} & a_{32}g_{32} & a_{33}g_{33} & a_{34}g_{34} \\ a_{41}g_{41} & a_{42}g_{42} & a_{43}g_{43} & a_{44}g_{44} \end{pmatrix}\mathbf{E}_{AB} + t_l m_{3c}\,\mathbf{E}_{AB}^\top\,Q_{4r}\,\mathbf{E}_{AB}\,. \tag{4.24}$$

Here the $4r\times r$ matrix $\mathbf{E}_{AB}$ is defined as

$$\mathbf{E}_{AB} := \widetilde{\Pi}_{r,r}\mathfrak{A}\mathcal{O} = \begin{pmatrix} -m_{3c}^{-1}\,\mathbf{A} \\ hm_{3c}^{-1}m_{4c}^{-1}\mathbf{B} \\ t_l^{-1}\mathbf{F}_1 \\ hm_{3c}^{-1}\mathbf{F}_2 \end{pmatrix},$$

where we have abbreviated that

$$\begin{aligned} &\mathbf{A} := \left(I_r+\Sigma_a^2\right)^{-1/2}\mathcal{O}, \quad \mathbf{B} := \left(1+\Sigma_b^2\right)^{-1/2}\widehat{\Sigma}_b\mathcal{M}_r^\top\widehat{\Sigma}_a\mathcal{O}, \\ &\mathbf{F}_1 := t_l\mathbf{V}_a\widehat{\Sigma}_a\mathcal{O}-\mathbf{V}_b\widehat{\Sigma}_b^2\mathcal{M}_r^\top\widehat{\Sigma}_a\mathcal{O}, \quad \mathbf{F}_2 := \mathbf{V}_a\widehat{\Sigma}_a\mathcal{O}-\mathbf{V}_b\widehat{\Sigma}_b^2\mathcal{M}_r^\top\widehat{\Sigma}_a\mathcal{O}. \end{aligned} \tag{4.25}$$

In the above derivation, we also used the identities $f_c(\theta_l) = m_{3c}(\theta_l)m_{4c}(\theta_l)/h^2(\theta_l) = t_l$. Expanding (4.24), we obtain that

$$\begin{aligned}
&\sqrt{n}\mathcal{O}^\top \mathcal{E}_r^{(2)} \mathcal{O} \\
&\overset{d}{\sim} t_l\, \mathbf{A}^\top \left[\frac{a_{11}}{m_{3c}} g_{11} + \sqrt{n}\mathbb{IE} \left(\mathbf{O}_1^\top X_1 S + \widetilde{\mathbf{W}}_1^\top X_R^{(1)} \right) \left(S^\top X_1^\top \mathbf{O}_1 + (X_R^{(1)})^\top \widetilde{\mathbf{W}}_1 \right) \right] \mathbf{A} \\
&- \left\{ \mathbf{A}^\top \left[\frac{a_{12}}{h} g_{12} + \sqrt{n} \left(\mathbf{O}_1^\top X_1 S + \widetilde{\mathbf{W}}_1^\top X_R^{(1)} \right) \left(S^\top Y_1^\top \mathbf{O}_2 + (Y_R^{(1)})^\top \widetilde{\mathbf{W}}_2 \right) \right] \mathbf{B} + c.t. \right\} \\
&+ \mathbf{B}^\top \left[\frac{a_{22}}{m_{4c}} g_{22} + \sqrt{n}\mathbb{IE} \left(\mathbf{O}_2^\top Y_1 S + \widetilde{\mathbf{W}}_2^\top Y_R^{(1)} \right) \left(S^\top Y_1^\top \mathbf{O}_2 + (Y_R^{(1)})^\top \widetilde{\mathbf{W}}_2 \right) \right] \mathbf{B} \\
&- \left[\mathbf{A}^\top \left(a_{13} g_{13} - \sqrt{n}\, \mathbf{O}_1^\top X_1 V - \sqrt{n} \widetilde{\mathbf{W}}_1^\top X_{RL}^{(1)} \right) \mathbf{F}_1 + c.t. \right] - \left[\mathbf{A}^\top \left(\frac{m_{4c}}{h} a_{14} g_{14} \right) \mathbf{F}_2 + c.t. \right] \\
&+ \left[\mathbf{B}^\top \left(\frac{h}{m_{4c}} a_{23} g_{23} \right) \mathbf{F}_1 + c.t. \right] + \left[\mathbf{B}^\top \left(a_{24} g_{24} - \sqrt{n}\, \mathbf{O}_2^\top Y_1 V - \sqrt{n} \widetilde{\mathbf{W}}_2^\top Y_{RL}^{(1)} \right) \mathbf{F}_2 + c.t. \right] \\
&+ t_l^{-1} \mathbf{F}_1^\top \left(m_{3c} a_{33} g_{33} - \sqrt{n}\mathbb{IE} (X_{RL}^{(2)})^\top X_{RL}^{(2)} \right) \mathbf{F}_1 + \mathbf{F}_2^\top \left(m_{4c} a_{44} g_{44} - \sqrt{n}\mathbb{IE} (Y_{RL}^{(2)})^\top Y_{RL}^{(2)} \right) \mathbf{F}_2 \\
&+ \left[\mathbf{F}_1^\top \left(h a_{34} g_{34} \right) \mathbf{F}_2 + c.t. \right],
\end{aligned} \tag{4.26}$$

where "*c.t.*" means the (conjugate) transpose of the preceding term.

In (4.26), $\sqrt{n}X_{RL}^{(1)}$ and $\sqrt{n}Y_{RL}^{(1)}$ have i.i.d. Gaussian entries of mean 0 and variance 1, and are independent of all the other terms. So we rename them as two n-independent Gaussian matrices

$$\widetilde{g}_{13} := -\sqrt{n}X_{RL}^{(1)}, \quad \widetilde{g}_{24} := -\sqrt{n}Y_{RL}^{(1)}. \tag{4.27}$$

Moreover, the matrices $\sqrt{n}\mathbb{IE}(X_{RL}^{(2)})^\top X_{RL}^{(2)}$ and $\sqrt{n}\mathbb{IE}(Y_{RL}^{(2)})^\top Y_{RL}^{(2)}$ are also independent of all the other terms. With classical CLT, we obtain that

$$-\sqrt{n}\mathbb{IE}(X_{RL}^{(2)})^\top X_{RL}^{(2)} \overset{d}{\sim} \sqrt{c_1}\widetilde{g}_{33}, \quad -\sqrt{n}\mathbb{IE}(Y_{RL}^{(2)})^\top Y_{RL}^{(2)} \overset{d}{\sim} \sqrt{c_2}\widetilde{g}_{44}, \tag{4.28}$$

where $\widetilde{g}_{33}$ and $\widetilde{g}_{44}$ are $r \times r$ symmetric GOE with entries $(\widetilde{g}_{33})_{ij} \sim \mathcal{N}(0, 1+\delta_{ij})$ and $(\widetilde{g}_{44})_{ij} \sim \mathcal{N}(0, 1+\delta_{ij})$. It remains to show the CLT for the following matrix

$$\begin{aligned}
\Theta : = & \, t_l\, \mathbf{A}^\top \left[\sqrt{n}\mathbb{IE} \left(\mathbf{O}_1^\top X_1 S + \widetilde{\mathbf{W}}_1^\top X_R^{(1)} \right) \left(S^\top X_1^\top \mathbf{O}_1 + (X_R^{(1)})^\top \widetilde{\mathbf{W}}_1 \right) \right] \mathbf{A} \\
&- \left\{ \mathbf{A}^\top \left[\sqrt{n} \left(\mathbf{O}_1^\top X_1 S + \widetilde{\mathbf{W}}_1^\top X_R^{(1)} \right) \left(S^\top Y_1^\top \mathbf{O}_2 + (Y_R^{(1)})^\top \widetilde{\mathbf{W}}_2 \right) \right] \mathbf{B} + c.t. \right\} \\
&+ \mathbf{B}^\top \left[\sqrt{n}\mathbb{IE} \left(\mathbf{O}_2^\top Y_1 S + \widetilde{\mathbf{W}}_2^\top Y_R^{(1)} \right) \left(S^\top Y_1^\top \mathbf{O}_2 + (Y_R^{(1)})^\top \widetilde{\mathbf{W}}_2 \right) \right] \mathbf{B} \\
&+ \left[\mathbf{A}^\top \left(\sqrt{n}\, \mathbf{O}_1^\top X_1 V \right) \mathbf{F}_1 + c.t. \right] - \left[\mathbf{B}^\top \left(\sqrt{n}\, \mathbf{O}_2^\top Y_1 V \right) \mathbf{F}_2 + c.t. \right].
\end{aligned} \tag{4.29}$$

We decompose Θ into the sum of four matrices,

$$\Theta := \Theta_1 + \Theta_2 + \Theta_3 + \Theta_4,$$

where

$$\begin{aligned}
\Theta_1 : = & \, \mathbf{B}^\top \left[\sqrt{n}\mathbb{IE} \widetilde{\mathbf{W}}_2^\top Y_R^{(1)} (Y_R^{(1)})^\top \widetilde{\mathbf{W}}_2 + \sqrt{n} \left(\mathbf{O}_2^\top Y_1 S (Y_R^{(1)})^\top \widetilde{\mathbf{W}}_2 + c.t. \right) \right] \mathbf{B} \\
&- \left[\mathbf{A}^\top \left(\sqrt{n} \left(\mathbf{O}_1^\top X_1 S + \widetilde{\mathbf{W}}_1^\top X_R^{(1)} \right) (Y_R^{(1)})^\top \widetilde{\mathbf{W}}_2 \right) \mathbf{B} + c.t. \right],
\end{aligned}$$

$$\begin{aligned}
\Theta_2 : = & \, t_l \, \mathbf{A}^\top \left[\sqrt{n}\mathbb{I}\mathbb{E}\widetilde{\mathbf{W}}_1^\top X_R^{(1)} (X_R^{(1)})^\top \widetilde{\mathbf{W}}_1 + \sqrt{n} \left(\mathbf{O}_1^\top \, X_1 S (X_R^{(1)})^\top \widetilde{\mathbf{W}}_1 + c.t.\right)\right] \mathbf{A} \\
& - \left[\mathbf{B}^\top \left(\sqrt{n} \, \mathbf{O}_2^\top \, Y_1 S (X_R^{(1)})^\top \widetilde{\mathbf{W}}_1\right) \mathbf{A} + c.t.\right], \\
\Theta_3 : = & \left[\mathbf{A}^\top \left(\sqrt{n} \, \mathbf{O}_1^\top \, X_1 V\right) \mathbf{F}_1 + c.t.\right] - \left[\mathbf{A}^\top \left(\sqrt{n} \, \mathbf{O}_1^\top \, X_1 S S^\top Y_1^\top \, \mathbf{O}_2\right) \mathbf{B} + c.t.\right] \\
& + t_l \, \mathbf{A}^\top \left[\sqrt{n}\mathbb{I}\mathbb{E}(\mathbf{O}_1^\top \, X_1 S S^\top X_1^\top \, \mathbf{O}_1)\right] \mathbf{A}, \\
\Theta_4 : = & - \left[\mathbf{B}^\top \left(\sqrt{n} \, \mathbf{O}_2^\top \, Y_1 V\right) \mathbf{F}_2 + c.t.\right] + \mathbf{B}^\top \left[\sqrt{n}\mathbb{I}\mathbb{E}\left(\mathbf{O}_2^\top \, Y_1 S S^\top Y_1^\top \, \mathbf{O}_2\right)\right] \mathbf{B}.
\end{aligned}$$

The decomposition is chosen as follows: Θ_1 contains all the terms that depend on $Y_R^{(1)}$, Θ_2 contains all the remaining terms that depend on $X_R^{(1)}$, Θ_3 contains all the remaining terms that depend on X_1, and Θ_4 contains all the remaining terms that depend on Y_1. Using (2.8) and Lemma 4.3, we can obtain the following large deviation estimates as in (4.17): for any small constant $\varepsilon > 0$,

$$\|X_1 V\|_{\max} + \|X_1\|_{\max} \leqslant n^{-1/2+\varepsilon}, \quad \|X_1 X_1^\top - I_\rho\|_{\max} \leqslant n^{-1/2+\varepsilon}, \tag{4.30}$$

$$\|Y_1 V\|_{\max} + \|Y_1\|_{\max} \leqslant n^{-1/2+\varepsilon}, \quad \|Y_1 Y_1^\top - I_\rho\|_{\max} \leqslant n^{-1/2+\varepsilon}, \tag{4.31}$$

with high probability. Combining (4.30) and (4.31) with the facts $SS^\top = I_n - VV^\top$ and $\rho = \mathrm{O}(n^{2\tau_0})$, we can simplify Θ_3 and Θ_4 as

$$\Theta_\alpha = \Theta_\alpha' + \mathrm{O}_\prec(n^{-1/2+4\tau_0}), \quad \alpha = 3, 4,$$

where

$$\begin{aligned}
\Theta_3' := & \left[\mathbf{A}^\top \left(\sqrt{n} \, \mathbf{O}_1^\top \, X_1 V\right) \mathbf{F}_1 + c.t.\right] - \left[\mathbf{A}^\top \left(\sqrt{n} \, \mathbf{O}_1^\top \, X_1 Y_1^\top \, \mathbf{O}_2\right) \mathbf{B} + c.t.\right] \\
& + t_l \, \mathbf{A}^\top \left[\sqrt{n}\mathbb{I}\mathbb{E}\left(\mathbf{O}_1^\top \, X_1 X_1^\top \, \mathbf{O}_1\right)\right] \mathbf{A}, \\
\Theta_4' := & - \left[\mathbf{B}^\top \left(\sqrt{n} \, \mathbf{O}_2^\top \, Y_1 V\right) \mathbf{F}_2 + c.t.\right] + \mathbf{B}^\top \left[\sqrt{n}\mathbb{I}\mathbb{E}\left(\mathbf{O}_2^\top \, Y_1 Y_1^\top \, \mathbf{O}_2\right)\right] \mathbf{B}.
\end{aligned}$$

Now we show that the four terms Θ_1, Θ_2, Θ_3', and Θ_4' are all asymptotically Gaussian. The proof of the following lemma is standard, and for the reader's convenience, we give the details in Appendix B.

Lemma 4.6. *We have the following results:*

(i) conditioning on X_1, Y_1 and $X_R^{(1)}$ satisfying (4.17), Θ_1 is asymptotically Gaussian with zero mean;

(ii) conditioning on X_1 and Y_1 satisfying (4.17), Θ_2 is asymptotically Gaussian with zero mean;

(iii) conditioning on Y_1 satisfying (4.31) and V satisfying (4.5), Θ_3' is asymptotically Gaussian with zero mean;

(iv) conditioning on V satisfying (4.5), Θ_4' is asymptotically Gaussian with zero mean.

With Lemma 4.6, we obtain that Θ converges in distribution to a centered Gaussian matrix. It remains to determine the covariances of this matrix. First, we calculate the covariances for Θ_1. Note that conditioning on X_1, Y_1 and $X_R^{(1)}$ satisfying (4.17) and using $\widetilde{r} = \mathrm{O}(n^{2\tau_0})$, we have that

$$(\mathbf{B}^\top \, \mathbf{O}_2^\top \, Y_1 S S^\top Y_1^\top \, \mathbf{O}_2 \, \mathbf{B})_{ij} = (\mathbf{B}^\top \, \mathbf{O}_2^\top \, \mathbf{O}_2 \, \mathbf{B})_{ij} + \mathrm{O}(n^{-1/2+2\tau_0+\varepsilon}),$$

and

$$\left[\mathbf{A}^\top \left(\mathbf{O}_1^\top \, X_1 S + \widetilde{\mathbf{W}}_1^\top X_R^{(1)}\right) \left(S^\top X_1^\top \, \mathbf{O}_1 + (X_R^{(1)})^\top \widetilde{\mathbf{W}}_1\right) \mathbf{A}\right]_{ij}$$

$$= \left[\mathbf{A}^\top \left(\mathbf{O}_1^\top \mathbf{O}_1 + \widetilde{\mathbf{W}}_1^\top \widetilde{\mathbf{W}}_1\right) \mathbf{A}\right]_{ij} + \mathrm{O}(n^{-1/2+2\tau_0+\varepsilon}) = (\mathbf{A}^\top \mathbf{A})_{ij} + \mathrm{O}(n^{-1/2+2\tau_0+\varepsilon}).$$

With these two identities, we can calculate that

$$\begin{aligned}
&\mathbb{E}_{Y_R^{(1)}}(\Theta_1)_{ij}(\Theta_1)_{i'j'} = \left(\mathbf{B}^\top\mathbf{B}\right)_{ii'} (\mathbf{B}^\top \widetilde{\mathbf{W}}_2^\top \widetilde{\mathbf{W}}_2 \mathbf{B})_{jj'} + \left(\mathbf{B}^\top\mathbf{B}\right)_{ij'} (\mathbf{B}^\top \widetilde{\mathbf{W}}_2^\top \widetilde{\mathbf{W}}_2 \mathbf{B})_{ji'} \\
&+ (\mathbf{B}^\top \widetilde{\mathbf{W}}_2^\top \widetilde{\mathbf{W}}_2 \mathbf{B})_{ii'} (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{jj'} + (\mathbf{B}^\top \widetilde{\mathbf{W}}_2^\top \widetilde{\mathbf{W}}_2 \mathbf{B})_{ij'} (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ji'} \\
&+ (\mathbf{A}^\top \mathbf{A})_{ii'} (\mathbf{B}^\top \widetilde{\mathbf{W}}_2^\top \widetilde{\mathbf{W}}_2 \mathbf{B})_{jj'} + (\mathbf{A}^\top \mathbf{A})_{ij'} (\mathbf{B}^\top \widetilde{\mathbf{W}}_2^\top \widetilde{\mathbf{W}}_2 \mathbf{B})_{ji'} \\
&+ (\mathbf{B}^\top \widetilde{\mathbf{W}}_2^\top \widetilde{\mathbf{W}}_2 \mathbf{B})_{ii'} (\mathbf{A}^\top \mathbf{A})_{jj'} + (\mathbf{B}^\top \widetilde{\mathbf{W}}_2^\top \widetilde{\mathbf{W}}_2 \mathbf{B})_{ij'} (\mathbf{A}^\top \mathbf{A})_{ji'} + \mathrm{O}(n^{-1/2+2\tau_0+\varepsilon}).
\end{aligned}$$

Similarly, conditioning on X_1 and Y_1 satisfying (4.17), we can calculate the covariances for Θ_2 as

$$\begin{aligned}
&\mathbb{E}_{X_R^{(1)}}(\Theta_2)_{ij}(\Theta_2)_{i'j'} = t_l^2 (\mathbf{A}^\top \mathbf{A})_{ii'} (\mathbf{A}^\top \widetilde{\mathbf{W}}_1^\top \widetilde{\mathbf{W}}_1 \mathbf{A})_{jj'} + t_l^2 (\mathbf{A}^\top \mathbf{A})_{ij'} (\mathbf{A}^\top \widetilde{\mathbf{W}}_1^\top \widetilde{\mathbf{W}}_1 \mathbf{A})_{ji'} \\
&+ t_l^2 (\mathbf{A}^\top \widetilde{\mathbf{W}}_1^\top \widetilde{\mathbf{W}}_1 \mathbf{A})_{ii'} (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{jj'} + t_l^2 (\mathbf{A}^\top \widetilde{\mathbf{W}}_1^\top \widetilde{\mathbf{W}}_1 \mathbf{A})_{ij'} (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{ji'} \\
&+ (\mathbf{A}^\top \widetilde{\mathbf{W}}_1^\top \widetilde{\mathbf{W}}_1 \mathbf{A})_{ii'} (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{jj'} + (\mathbf{A}^\top \widetilde{\mathbf{W}}_1^\top \widetilde{\mathbf{W}}_1 \mathbf{A})_{ij'} (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ji'} \\
&+ (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ii'} (\mathbf{A}^\top \widetilde{\mathbf{W}}_1^\top \widetilde{\mathbf{W}}_1 \mathbf{A})_{jj'} + (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ij'} (\mathbf{A}^\top \widetilde{\mathbf{W}}_1^\top \widetilde{\mathbf{W}}_1 \mathbf{A})_{ji'} + \mathrm{O}(n^{-1/2+2\tau_0+\varepsilon}).
\end{aligned}$$

For the calculations of the covariances for Θ_3' and Θ_4', the entries of X_1 and Y_1 are not Gaussian anymore. Recall that their fourth moments $\mu_x^{(4)}$ and $\mu_y^{(4)}$ are defined in (2.27), and we denote their third moments as

$$\mu_x^{(3)} := n^{3/2}\mathbb{E}X_{11}^3, \quad \mu_y^{(3)} := n^{3/2}\mathbb{E}Y_{11}^3. \tag{4.32}$$

Then we can calculate the covariances for Θ_3' as

$$\begin{aligned}
&\mathbb{E}_{X_1}(\Theta_3')_{ij}(\Theta_3')_{i'j'} = \left(\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A}\right)_{ii'} \left[\left(\mathbf{F}_1^\top V^\top - \mathbf{B}^\top \mathbf{O}_2^\top Y_1\right)\left(V\mathbf{F}_1 - Y_1^\top \mathbf{O}_2 \mathbf{B}\right)\right]_{jj'} \\
&+ \left(\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A}\right)_{ij'} \left[\left(\mathbf{F}_1^\top V^\top - \mathbf{B}^\top \mathbf{O}_2^\top Y_1\right)\left(V\mathbf{F}_1 - Y_1^\top \mathbf{O}_2 \mathbf{B}\right)\right]_{ji'} \\
&+ \left[\left(\mathbf{F}_1^\top V^\top - \mathbf{B}^\top \mathbf{O}_2^\top Y_1\right)\left(V\mathbf{F}_1 - Y_1^\top \mathbf{O}_2 \mathbf{B}\right)\right]_{ii'} (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{jj'} \\
&+ \left[\left(\mathbf{F}_1^\top V^\top - \mathbf{B}^\top \mathbf{O}_2^\top Y_1\right)\left(V\mathbf{F}_1 - Y_1^\top \mathbf{O}_2 \mathbf{B}\right)\right]_{ij'} (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{ji'} \\
&+ t_l^2 \left(\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A}\right)_{ii'} \left(\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A}\right)_{jj'} + t_l^2 \left(\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A}\right)_{ij'} \left(\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A}\right)_{ji'} \\
&+ t_l^2 (\mu_x^{(4)} - 3) \sum_{1 \leqslant k \leqslant \rho} (\mathbf{O}_1 \mathbf{A})_{ki} (\mathbf{O}_1 \mathbf{A})_{ki'} (\mathbf{O}_1 \mathbf{A})_{kj} (\mathbf{O}_1 \mathbf{A})_{kj'} \\
&+ t_l \mu_x^{(3)} \frac{1}{\sqrt{n}} \sum_{1 \leqslant k \leqslant \rho, \mu \in \mathcal{I}_3} (\mathbf{O}_1 \mathbf{A})_{ki} (\mathbf{O}_1 \mathbf{A})_{ki'} (\mathbf{O}_1 \mathbf{A})_{kj'} \left(V\mathbf{F}_1 - Y_1^\top \mathbf{O}_2 \mathbf{B}\right)_{\mu j} \\
&+ t_l \mu_x^{(3)} \frac{1}{\sqrt{n}} \sum_{1 \leqslant k \leqslant \rho, \mu \in \mathcal{I}_3} (\mathbf{O}_1 \mathbf{A})_{kj} (\mathbf{O}_1 \mathbf{A})_{ki'} (\mathbf{O}_1 \mathbf{A})_{kj'} \left(V\mathbf{F}_1 - Y_1^\top \mathbf{O}_2 \mathbf{B}\right)_{\mu i} \\
&+ t_l \mu_x^{(3)} \frac{1}{\sqrt{n}} \sum_{1 \leqslant k \leqslant \rho, \mu \in \mathcal{I}_3} (\mathbf{O}_1 \mathbf{A})_{ki} (\mathbf{O}_1 \mathbf{A})_{kj} (\mathbf{O}_1 \mathbf{A})_{ki'} \left(V\mathbf{F}_1 - Y_1^\top \mathbf{O}_2 \mathbf{B}\right)_{\mu j'} \\
&+ t_l \mu_x^{(3)} \frac{1}{\sqrt{n}} \sum_{1 \leqslant k \leqslant \rho, \mu \in \mathcal{I}_3} (\mathbf{O}_1 \mathbf{A})_{ki} (\mathbf{O}_1 \mathbf{A})_{kj} (\mathbf{O}_1 \mathbf{A})_{kj'} \left(V\mathbf{F}_1 - Y_1^\top \mathbf{O}_2 \mathbf{B}\right)_{\mu i'} .
\end{aligned} \tag{4.33}$$

Using Lemma 4.3, we can check that

$$\|Y_1 \mathbf{e}\|_{\max} \prec n^{-1/2}, \quad \text{for} \quad \mathbf{e} := n^{-1/2}(1, 1, \cdots, 1)^\top \in \mathbb{R}^n.$$

Applying this estiamate and (4.31), we obtain that

$$(\mathbf{F}_1^\top V^\top - \mathbf{B}^\top \mathbf{O}_2^\top Y_1)(V\mathbf{F}_1 - Y_1^\top \mathbf{O}_2 \mathbf{B}) = \mathbf{F}_1^\top \mathbf{F}_1 + \mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B} + \mathrm{O}_\prec(n^{-1/2+2\tau_0}),$$

and that for any $1 \leqslant i \leqslant r$,

$$\frac{1}{\sqrt{n}} \sum_{\mu \in \mathcal{I}_3} \left(Y_1^\top \mathbf{O}_2 \mathbf{B}\right)_{\mu i} = (\mathbf{e}^\top Y_1^\top \mathbf{O}_2 \mathbf{B})_i = \mathrm{O}_\prec(n^{-1/2+2\tau_0}).$$

On the other hand, using (3.32) and (4.4), we obtain that

$$\|V^\top \mathbf{e}\|_{\max} \leqslant \|Z\mathbf{e}\|_{\max} + \|(V^\top - Z)\mathbf{e}\|_{\max} \prec n^{-1/2}. \tag{4.34}$$

Hence we also have that for any $1 \leqslant i \leqslant r$,

$$\frac{1}{\sqrt{n}} \sum_{\mu \in \mathcal{I}_3} (V\mathbf{F}_1)_{\mu i} = \mathrm{O}_\prec(n^{-1/2}).$$

The above calculations show that the $\mu_x^{(3)}$ terms are all negligible. For the $\mu_x^{(4)}$ term, by the assumptions of Proposition 4.1, we have that $\|\mathbf{W}_1\|_{\max} \leqslant n^{-\tau_0}$, which gives $(\mathbf{W}_1 \mathbf{A})_{ki} \lesssim n^{-\tau_0}$ for any k. With this fact, we obtain that

$$\sum_{\rho+1 \leqslant k \leqslant p} (\mathbf{W}_1 \mathbf{A})_{ki} (\mathbf{W}_1 \mathbf{A})_{ki'} (\mathbf{W}_1 \mathbf{A})_{kj} (\mathbf{W}_1 \mathbf{A})_{kj'} \lesssim n^{-2\tau_0} \sum_{\rho+1 \leqslant k \leqslant p} (\mathbf{W}_1 \mathbf{A})_{ki} (\mathbf{W}_1 \mathbf{A})_{ki'} \lesssim n^{-2\tau_0}.$$

Thus we can replace $\mathbf{O}_1 \mathbf{A}$ with $\mathbf{U}_a \mathbf{A}$ in $\sum_{1 \leqslant k \leqslant \rho} (\mathbf{O}_1 \mathbf{A})_{ki} (\mathbf{O}_1 \mathbf{A})_{ki'} (\mathbf{O}_1 \mathbf{A})_{kj} (\mathbf{O}_1 \mathbf{A})_{kj'}$ up to a negligible error. Collecting the above estimates, we can simplify (4.33) as

$$\begin{aligned}
\mathbb{E}_{X_1} (\Theta_3')_{ij} (\Theta_3')_{i'j'} = {} & (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{ii'} (\mathbf{F}_1^\top \mathbf{F}_1 + \mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{jj'} + (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{ij'} (\mathbf{F}_1^\top \mathbf{F}_1 + \mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ji'} \\
& + (\mathbf{F}_1^\top \mathbf{F}_1 + \mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ii'} (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{jj'} + (\mathbf{F}_1^\top \mathbf{F}_1 + \mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ij'} (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{ji'} \\
& + t_l^2 (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{ii'} (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{jj'} + t_l^2 (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{ij'} (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{ji'} \\
& + t_l^2 (\mu_x^{(4)} - 3) \sum_{k \in \mathcal{I}_1} \mathcal{U}_{ki} \mathcal{U}_{ki'} \mathcal{U}_{kj} \mathcal{U}_{kj'} + \mathrm{O}(n^{-2\tau_0})
\end{aligned}$$

with high probability, where we recall the notations in (2.25) and (4.25). With similar calculations, we can obtain the covariances of Θ_4' as with high probability,

$$\begin{aligned}
\mathbb{E}_{Y_1} (\Theta_4')_{ij} (\Theta_4')_{i'j'} = {} & (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ii'} (\mathbf{F}_2^\top \mathbf{F}_2)_{jj'} + (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ij'} (\mathbf{F}_2^\top \mathbf{F}_2)_{ji'} \\
& + (\mathbf{F}_2^\top \mathbf{F}_2)_{ii'} (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{jj'} + \left(\mathbf{F}_2^\top \mathbf{F}_2\right)_{ij'} (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ji'} \\
& + (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ii'} (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{jj'} + (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ij'} (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ji'} \\
& + (\mu_y^{(4)} - 3) \sum_{k \in \mathcal{I}_2} (\mathbf{U}_b \mathbf{B})_{ki} (\mathbf{U}_b \mathbf{B})_{ki'} (\mathbf{U}_b \mathbf{B})_{kj} (\mathbf{U}_b \mathbf{B})_{kj'} + \mathrm{O}(n^{-2\tau_0}).
\end{aligned}$$

Combining all the above calculations, we have shown that $\Theta = \Theta_1 + \Theta_2 + \Theta_3 + \Theta_4$ converges weakly to a centered Gaussian random matrix g_Θ with covariances

$$\mathbb{E}(g_\Theta)_{ij} (g_\Theta)_{i'j'} = t_l^2 (\mathbf{A}^\top \mathbf{A})_{ii'} (\mathbf{A}^\top \mathbf{A})_{jj'} + t_l^2 (\mathbf{A}^\top \mathbf{A})_{ij'} (\mathbf{A}^\top \mathbf{A})_{ji'} + (\mathbf{B}^\top \mathbf{B})_{ii'} (\mathbf{B}^\top \mathbf{B})_{jj'} + (\mathbf{B}^\top \mathbf{B})_{ij'} (\mathbf{B}^\top \mathbf{B})_{ji'}$$

$$
\begin{aligned}
&+ (\mathbf{A}^\top \mathbf{A})_{ii'}(\mathbf{B}^\top\mathbf{B})_{jj'} + (\mathbf{A}^\top \mathbf{A})_{ij'}(\mathbf{B}^\top\mathbf{B})_{ji'} + (\mathbf{B}^\top\mathbf{B})_{ii'}(\mathbf{A}^\top \mathbf{A})_{jj'} + (\mathbf{B}^\top\mathbf{B})_{ij'}(\mathbf{A}^\top \mathbf{A})_{ji'} \\
&+ (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{ii'} \left(\mathbf{F}_1^\top\mathbf{F}_1\right)_{jj'} + (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{ij'} \left(\mathbf{F}_1^\top\mathbf{F}_1\right)_{ji'} \\
&+ \left(\mathbf{F}_1^\top\mathbf{F}_1\right)_{ii'} (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{jj'} + \left(\mathbf{F}_1^\top\mathbf{F}_1\right)_{ij'} (\mathbf{A}^\top \mathbf{O}_1^\top \mathbf{O}_1 \mathbf{A})_{ji'} \\
&+ (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ii'}(\mathbf{F}_2^\top\mathbf{F}_2)_{jj'} + (\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ij'}(\mathbf{F}_2^\top\mathbf{F}_2)_{ji'} \\
&+ (\mathbf{F}_2^\top\mathbf{F}_2)_{ii'}(\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{jj'} + (\mathbf{F}_2^\top\mathbf{F}_2)_{ij'}(\mathbf{B}^\top \mathbf{O}_2^\top \mathbf{O}_2 \mathbf{B})_{ji'} + t_l^2(\mu_x^{(4)} - 3) \sum_{k\in\mathcal{I}_1} \mathcal{U}_{ki}\mathcal{U}_{ki'}\mathcal{U}_{kj}\mathcal{U}_{kj'} \\
&+ (\mu_y^{(4)} - 3) \sum_{k\in\mathcal{I}_2} (\mathbf{U}_b\mathbf{B})_{ki} (\mathbf{U}_b\mathbf{B})_{ki'} (\mathbf{U}_b\mathbf{B})_{kj} (\mathbf{U}_b\mathbf{B})_{kj'} . \qquad (4.35)
\end{aligned}
$$

Notice that for any $i \in \gamma(l)$, we have that

$$
(\mathbf{U}_b\mathbf{B})_{ki} = \left[\sqrt{t_l} + \mathrm{O}(n^{-1/2+\delta})\right] \mathcal{V}_{ki}, \qquad (4.36)
$$

where we used the SVD (2.23) and the fact $t_i = t_l + \mathrm{O}(n^{-1/2+\delta})$ for any $i \in \gamma(l)$ by Definition 2.2. Hence the last term in (4.35) can be replaced by

$$
(\mu_y^{(4)} - 3) \sum_{k\in\mathcal{I}_2} \mathcal{V}_{ki}\mathcal{V}_{ki'}\mathcal{V}_{kj}\mathcal{V}_{kj'}, \quad \text{for} \quad i, j, i', j' \in \gamma(l).
$$

4.5 Step 5: Concluding the proof

Finally, combing (4.26), (4.27), (4.28) and (4.35), after a lengthy but straightforward calculation we obtain that $(\sqrt{n}\mathcal{O}^\top\mathcal{E}_r^{(2)}\mathcal{O})_{[\![\gamma(l)]\!]}$ coverages weakly to an $r \times r$ centered Gaussian matrix $\Upsilon_l^{(2)}$ with covariances

$$
\begin{aligned}
&\mathbb{E}(\Upsilon_l^{(2)})_{ij}(\Upsilon_l^{(2)})_{i'j'} = t_l^2 \left(\frac{a_c^2 + c_1}{1 - c_1} + \frac{a_c^2}{c_1} + 1\right) [(1-\mathcal{A})_{ii'}(1-\mathcal{A})_{jj'} + (1-\mathcal{A})_{ij'}(1-\mathcal{A})_{ji'}] \\
&+ \left(\frac{a_c^2 + c_2}{1 - c_2} + \frac{a_c^2}{c_2} + 1\right) (\mathcal{B}_{ii'}\mathcal{B}_{jj'} + \mathcal{B}_{ij'}\mathcal{B}_{ji'}) \\
&+ \left(\frac{a_c^2}{c_2}t_l^2 + \frac{a_c^2 + c_2}{1 - c_2} + 1\right) [(1-\mathcal{A})_{ii'}\mathcal{B}_{jj'} + \mathcal{B}_{ii'}(1-\mathcal{A})_{jj'} + (1-\mathcal{A})_{ij'}\mathcal{B}_{ji'} + \mathcal{B}_{ij'}(1-\mathcal{A})_{ji'}] \\
&+ \left(\frac{a_c^2 + c_1}{1 - c_1} + 1\right) [(1-\mathcal{A})_{ii'}(\mathcal{F}_1)_{jj'} + (\mathcal{F}_1)_{ii'}(1-\mathcal{A})_{jj'} + (1-\mathcal{A})_{ij'}(\mathcal{F}_1)_{ji'} + (\mathcal{F}_1)_{ij'}(1-\mathcal{A})_{ji'}] \\
&+ \frac{a_c^2}{c_1}t_l^2 [(1-\mathcal{A})_{ii'}(\mathcal{F}_2)_{jj'} + (\mathcal{F}_2)_{ii'}(1-\mathcal{A})_{jj'} + (1-\mathcal{A})_{ij'}(\mathcal{F}_2)_{ji'} + (\mathcal{F}_2)_{ij'}(1-\mathcal{A})_{ji'}] \\
&+ \frac{a_c^2}{c_2} [\mathcal{B}_{ii'}(\mathcal{F}_1)_{jj'} + (\mathcal{F}_1)_{ii'}\mathcal{B}_{jj'} + \mathcal{B}_{ij'}(\mathcal{F}_1)_{ji'} + (\mathcal{F}_1)_{ij'}\mathcal{B}_{ji'}] \\
&+ \left(\frac{a_c^2 + c_2}{1 - c_2} + 1\right) [\mathcal{B}_{ii'}(\mathcal{F}_2)_{jj'} + (\mathcal{F}_2)_{ii'}\mathcal{B}_{jj'} + \mathcal{B}_{ij'}(\mathcal{F}_2)_{ji'} + (\mathcal{F}_2)_{ij'}\mathcal{B}_{ji'}] \\
&+ t_l^{-2} \left(c_1\frac{a_c^2 + c_1}{1 - c_1} + c_1\right) [(\mathcal{F}_1)_{ii'}(\mathcal{F}_1)_{jj'} + (\mathcal{F}_1)_{ij'}(\mathcal{F}_1)_{ji'}] + \left(c_2\frac{a_c^2 + c_2}{1 - c_2} + c_2\right) [(\mathcal{F}_2)_{ii'}(\mathcal{F}_2)_{jj'} + (\mathcal{F}_2)_{ij'}(\mathcal{F}_2)_{ji'}] \\
&+ a_c^2 [(\mathcal{F}_1)_{ii'}(\mathcal{F}_2)_{jj'} + (\mathcal{F}_2)_{ii'}(\mathcal{F}_1)_{jj'} + (\mathcal{F}_1)_{ij'}(\mathcal{F}_2)_{ji'} + (\mathcal{F}_2)_{ij'}(\mathcal{F}_1)_{ji'}] \\
&+ t_l^2(\mu_x^{(4)} - 3) \sum_k \mathcal{U}_{ki}\mathcal{U}_{ki'}\mathcal{U}_{kj}\mathcal{U}_{kj'} + (\mu_y^{(4)} - 3) \sum_k \mathcal{V}_{ki}\mathcal{V}_{ki'}\mathcal{V}_{kj}\mathcal{V}_{kj'}.
\end{aligned}
$$

Here for simplicity, we introduced the following notations:

$$\begin{aligned}&\mathcal{A} := 1 - \mathbf{A}^\top \mathbf{A} = \mathcal{O}^\top \widehat{\Sigma}_a^2 \mathcal{O}, \quad \mathcal{B} := \mathbf{B}^\top \mathbf{B} = \mathcal{O}^\top \widehat{\Sigma}_a \mathcal{M}_r \widehat{\Sigma}_b \left(1 + \Sigma_b^2\right)^{-1} \widehat{\Sigma}_b \mathcal{M}_r^\top \widehat{\Sigma}_a \mathcal{O}, \\ &\mathcal{F}_1 := \mathbf{F}_1^\top \mathbf{F}_1 = t_l^2 \mathcal{A} + (1-2t_l)\mathcal{C} - \mathcal{B}, \quad \mathcal{F}_2 := \mathbf{F}_2^\top \mathbf{F}_2 = \mathcal{A} - \mathcal{C} - \mathcal{B},\end{aligned} \tag{4.37}$$

where we abbreviated that

$$\mathcal{C} := \mathcal{O}^\top \widehat{\Sigma}_a \mathcal{M}_r \widehat{\Sigma}_b^2 \mathcal{M}_r^\top \widehat{\Sigma}_a \mathcal{O} = \operatorname{diag}(t_1, \cdots, t_r).$$

Now we plug (4.37) into $\mathbb{E}(\Upsilon_l^{(2)})_{ij}(\Upsilon_l^{(2)})_{i'j'}$ and simplify the resulting expression. After a tedious but straightforward calculation, we can show that

$$\begin{aligned}\mathbb{E}(\Upsilon_l^{(2)})_{ij}(\Upsilon_l^{(2)})_{i'j'} &= \delta_{ii'} \left[t_l^2 \frac{a_c^2 + c_1}{c_1(1-c_1)} + \left(\frac{a_c^2+1}{1-c_1}(1-2t_l) - \frac{t_l^2 a_c^2}{c_1} \right) \mathcal{C} \right]_{jj'} \\ &+ \mathcal{C}_{ii'} \left[\left(\frac{a_c^2+1}{1-c_1}(1-2t_l) - \frac{t_l^2 a_c^2}{c_1} \right) + \left(\frac{(1-c_2)(1-2t_l)^2}{c_2} + \frac{(1-c_1)t_l^2}{c_1} - 2(1-2t_l) \right) a_c^2 \mathcal{C} \right]_{jj'} \\ &- (1-2t_l)\left(\mathcal{A}_{ii'}\mathcal{C}_{jj'} + \mathcal{C}_{ii'}\mathcal{A}_{jj'}\right) - \left(\mathcal{B}_{ii'}\mathcal{C}_{jj'} + \mathcal{C}_{ii'}\mathcal{B}_{jj'}\right) - t_l^2 \mathcal{A}_{ii'}\mathcal{A}_{jj'} - \mathcal{B}_{ii'}\mathcal{B}_{jj'} + \left(\mathcal{A}_{ii'}\mathcal{B}_{jj'} + \mathcal{B}_{ii'}\mathcal{A}_{jj'}\right) \\ &+ (i' \leftrightarrow j') + t_l^2(\mu_x^{(4)} - 3)\sum_k \mathcal{U}_{ki}\mathcal{U}_{ki'}\mathcal{U}_{kj}\mathcal{U}_{kj'} + t_l^2(\mu_y^{(4)} - 3)\sum_k \mathcal{V}_{ki}\mathcal{V}_{ki'}\mathcal{V}_{kj}\mathcal{V}_{kj'},\end{aligned}$$

where $(i' \leftrightarrow j')$ means an expression obtained by exchanging i' and j' in *all the preceding terms* (i.e. the terms in the first three lines).

On the other hand, using (3.43) and (3.45) we can check that $(\sqrt{n}\mathcal{O}^\top \mathcal{E}_r^{(1)} \mathcal{O})_{[\![\gamma(l)]\!]}$ converges weakly to an $r \times r$ centered Gaussian matrix $\Upsilon_l^{(1)}$ with covariances

$$\begin{aligned}\mathbb{E}(\Upsilon_l^{(1)})_{ij}(\Upsilon_l^{(1)})_{i'j'} &= (2t_l - 1)\mathcal{C}_{ii'}\mathcal{C}_{jj'} + t_l^2 \mathcal{A}_{ii'}\mathcal{A}_{jj'} + \mathcal{B}_{ii'}\mathcal{B}_{jj'} + (1-2t_l)\left(\mathcal{A}_{ii'}\mathcal{C}_{jj'} + \mathcal{C}_{ii'}\mathcal{A}_{jj'}\right) \\ &- \left(\mathcal{A}_{ii'}\mathcal{B}_{jj'} + \mathcal{B}_{ii'}\mathcal{A}_{jj'}\right) + \left(\mathcal{B}_{ii'}\mathcal{C}_{jj'} + \mathcal{C}_{ii'}\mathcal{B}_{jj'}\right) + (i' \leftrightarrow j') + (\mu_z^{(4)} - 3)\sum_k \mathcal{W}_{k,ij}\mathcal{W}_{k,i'j'},\end{aligned}$$

where we recall the notation in (2.24). Then by (4.8), we know that

$$\left(\sqrt{n}\mathcal{O}^\top \mathcal{E}_r(\theta_l)\mathcal{O}\right)_{[\![\gamma(l)]\!]} = \left(\sqrt{n}\mathcal{O}^\top \mathcal{E}_r^{(1)}\mathcal{O}\right)_{[\![\gamma(l)]\!]} + \left(\sqrt{n}\mathcal{O}^\top \mathcal{E}_r^{(2)}\mathcal{O}\right)_{[\![\gamma(l)]\!]}$$

converges weakly to a centered Gaussian matrix $\widetilde{\Upsilon}_l$ with covariances

$$\mathbb{E}(\widetilde{\Upsilon}_l)_{ij}(\widetilde{\Upsilon}_l)_{i'j'} = \mathbb{E}(\Upsilon_l^{(1)})_{ij}(\Upsilon_l^{(1)})_{i'j'} + \mathbb{E}(\Upsilon_l^{(2)})_{ij}(\Upsilon_l^{(2)})_{i'j'}.$$

Finally, using $\mathcal{C}_{jj'} = t_l \delta_{jj'} + \mathrm{O}(n^{-1/2+\delta})$ for $j, j' \in \gamma(l)$, we can check that the covariances of $\widetilde{\Upsilon}_l$ are asymptotically equal to (2.26). This concludes Proposition 4.1.

5 Proof of Theorem 2.3

Combining Proposition 4.1 with Proposition 3.11, we see that (2.31) holds in the almost Gaussian case. Hence to conclude Theorem 2.3, it suffices to show that the general case is sufficiently close to the almost Gaussian case regarding the outliers. In particular, by (3.41), (3.42) and (4.6), we only need to show that the asymptotic distribution of $\mathcal{M}(\theta_l)$ in (4.7) for general X and Y is the same as that of $\mathcal{M}^g(\theta_l)$ defined for almost Gaussian $X \equiv X^g$ and $Y \equiv Y^g$. Corresponding to (4.1) and (4.2), we define the index set (where "s" stands for "small")

$$\mathcal{I}_s := \Big\{1 \leqslant k \leqslant p : \max_{1\leqslant i\leqslant r} |\mathbf{u}_i^a(k)| \leqslant n^{-\tau_0}\Big\} \cup \Big\{1 \leqslant k - p \leqslant q : \max_{1\leqslant i\leqslant r} |\mathbf{u}_i^b(k)| \leqslant n^{-\tau_0}\Big\}.$$

Corresponding to (3.6) and (3.7), we define a new self-adjoint block matrix H^g and its resolvent as

$$H^g(z) := \begin{bmatrix} 0 & \begin{pmatrix} X^g & 0 \\ 0 & Y^g \end{pmatrix} \\ \begin{pmatrix} (X^g)^\top & 0 \\ 0 & (Y^g)^\top \end{pmatrix} & \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}^{-1} \end{bmatrix}, \quad G^g(z) := [H^g(z)]^{-1}.$$

Here X^g and Y^g are defined through

$$X^g_{i\mu} = \begin{cases} X_{i\mu}, & \text{if } i \notin \mathcal{I}_s \\ g^{(1)}_{i\mu}, & \text{if } i \in \mathcal{I}_s \end{cases}, \quad Y^g_{i\mu} = \begin{cases} Y_{i\mu}, & \text{if } i \notin \mathcal{I}_s \\ g^{(2)}_{i\mu}, & \text{if } i \in \mathcal{I}_s \end{cases}, \tag{5.1}$$

where $g^{(1)}_{i\mu}$ and $g^{(2)}_{i\mu}$ are i.i.d. Gaussian random variables independent of X and Y, and with mean zero and variance n^{-1}. Note that X^g and Y^g satisfy the setting of Proposition 4.1.

Define the index set

$$\mathcal{J}_s := \{(i,\mu) : i \in \mathcal{I}_1 \cap \mathcal{I}_s, \mu \in \mathcal{I}_3\} \cup \{(i,\mu) : i \in \mathcal{I}_2 \cap \mathcal{I}_s, \mu \in \mathcal{I}_4\}.$$

We choose a bijective ordering map Φ on $\mathcal{J}_s$ as

$$\Phi : \mathcal{J}_s \to \{1, \ldots, \gamma_{\max}\}, \quad \gamma_{\max} := |\mathcal{J}_s| = |\mathcal{I}_s| \cdot n.$$

For simplicity of notations, we abbreviate

$$W := \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}, \quad W^g := \begin{pmatrix} X^g & 0 \\ 0 & Y^g \end{pmatrix}. \tag{5.2}$$

For any $1 \leqslant \gamma \leqslant \gamma_{\max}$, we define the $(\mathcal{I}_1 \cup \mathcal{I}_2) \times (\mathcal{I}_3 \cup \mathcal{I}_4)$ matrix $W^{\{\gamma\}}$ such that

$$W^{\{\gamma\}}_{i\mu} = \begin{cases} W_{i\mu}, & \text{if } \Phi(i,\mu) \leqslant \gamma \\ W^g_{i\mu}, & \text{if } \Phi(i,\mu) > \gamma \end{cases}, \quad \text{and} \quad W^{\{\gamma\}}_{i\mu} = W_{i\mu} = W^g_{i\mu} \quad \text{for} \quad (i,\mu) \notin \mathcal{J}_s.$$

Correspondingly, we define

$$H^{\{\gamma\}}(z) := \begin{bmatrix} 0 & W^{\{\gamma\}} \\ (W^{\{\gamma\}})^\top & \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}^{-1} \end{bmatrix}, \quad G^{\{\gamma\}} := [H^{\{\gamma\}}(z)]^{-1}.$$

Note that we have $G^{\{0\}} = G^g$ and $G^{\{\gamma_{\max}\}} = G$. Now for $\Phi(i,\mu) = \gamma$, we can write that

$$H^{\{\gamma\}} = Q^{\{\gamma\}} + W_{i\mu}E^{\{\gamma\}}, \quad H^{\{\gamma-1\}} = Q^{\{\gamma\}} + W^g_{i\mu}E^{\{\gamma\}}, \tag{5.3}$$

where $E^{\{\gamma\}}$ is defined by

$$(E^{\{\gamma\}})_{ab} = \mathbf{1}_{(a,b)=(i,\mu)} + \mathbf{1}_{(a,b)=(\mu,i)}, \tag{5.4}$$

and $Q^{\{\gamma\}}$ is a random matrix with zero (i,μ)-th and (μ,i)-th entries. In particular, $\mathcal{Q}^{\{\gamma\}}$ is independent of $W_{i\mu}$ and $W^g_{i\mu}$. For simplicity of notations, for any γ we denote that

$$T^{\{\gamma\}} := G^{\{\gamma\}}, \quad S^{\{\gamma\}} := G^{\{\gamma-1\}}, \quad R^{\{\gamma\}} := (Q^{\{\gamma\}})^{-1}. \tag{5.5}$$

With this notation, given any function f, we can write that

$$\mathbb{E}f\left(G\right) - \mathbb{E}f\left(G^g\right) = \sum_{\gamma=1}^{\gamma_{\max}} \left[\mathbb{E}f\left(T^{\{\gamma\}}\right) - \mathbb{E}f\left(S^{\{\gamma\}}\right)\right]. \tag{5.6}$$

We will estimate each term in the sum using resolvent expansions. More precisely, by (5.3) we have that

$$T^{\{\gamma\}} = \left(Q^{\{\gamma\}} + W_{i\mu}E^{\{\gamma\}}\right)^{-1} = \left(1 + W_{i\mu}R^{\{\gamma\}}E^{\{\gamma\}}\right)^{-1} R^{\{\gamma\}}.$$

For any fixed $k \in \mathbb{N}$, we can expand $T^{\{\gamma\}}$ till order k as

$$T^{\{\gamma\}} = \sum_{s=0}^{k} (-W_{i\mu})^s \left(R^{\{\gamma\}}E^{\{\gamma\}}\right)^s R^{\{\gamma\}} + (-W_{i\mu})^{k+1} \left(R^{\{\gamma\}}E^{\{\gamma\}}\right)^{k+1} T^{\{\gamma\}}. \tag{5.7}$$

On the other hand, we can also expand $R^{\{\gamma\}}$ in terms of $T^{\{\gamma\}}$ as

$$R^{\{\gamma\}} = \left(1 - W_{i\mu}T^{\{\gamma\}}E^{\{\gamma\}}\right)^{-1} T^{\{\gamma\}} = \sum_{s=0}^{k} W_{i\mu}^s \left(T^{\{\gamma\}}E^{\{\gamma\}}\right)^s T^{\{\gamma\}} + W_{i\mu}^{k+1} \left(T^{\{\gamma\}}E^{\{\gamma\}}\right)^{k+1} R^{\{\gamma\}}. \tag{5.8}$$

We can get similar expansions for $S^{\{\gamma\}}$ and $R^{\{\gamma\}}$ by replacing $(T^{\{\gamma\}}, W_{i\mu})$ with $(S^{\{\gamma\}}, W^g_{i\mu})$. We will combine these resolvent expansions with the Taylor expansion of f to estimate the right-hand side of (5.6). Before doing that, we first introduced the concept of regularized resolvents in order to avoid possible singular behaviours of the resolvents $G^{\{\gamma\}}$ on exceptional low-probability events.

Definition 5.1 (Regularized resolvents)**.** *For $z = E + \mathrm{i}\eta \in \mathbb{C}_+$, we define the regularized resolvent $\widehat{G}(z)$ as*

$$\widehat{G}(z) := \left[H(z) - zn^{-10}\begin{pmatrix} I_{p+q} & 0 \\ 0 & 0 \end{pmatrix}\right]^{-1}.$$

We can define $\widehat{G}^g$ and $\widehat{G}^{\{\gamma\}}$ in a similar way.

The main reason for introducing the regularized resolvents is that they satisfy the deterministic bounds:

$$\|\widehat{G}(z)\| + \|\widehat{G}^g(z)\| + \max_\gamma \|\widehat{G}^{\{\gamma\}}(z)\| \lesssim n^{10}\eta^{-1}, \quad \text{for} \quad \eta = \operatorname{Im} z. \tag{5.9}$$

This estimate has been proved in Lemma 3.6 of [40]. In particular, if we choose $\eta \geqslant n^{-C}$ for a constant $C > 0$, then (5.9) justfies the assumption of Lemma 3.2 (iii), which will be used in the proof when we bound expectations of polynomials of regularized resolvent entries.

With a standard perturbation argument, we can easily control the difference between $\widehat{G}(z)$ and $G(z)$.

Claim 5.2. *Suppose there exists a high probability event Ξ on which $\|G(z)\|_{\max} = \mathrm{O}(1)$ for z belonging to some subset. Then we have that*

$$\|G(z) - \widehat{G}(z)\|_{\max} \leqslant n^{-8} \quad \text{on} \quad \Xi. \tag{5.10}$$

Proof. For $t \in [0,1]$, we define

$$G_t(z) := \left[H(z) - tzn^{-10} \begin{pmatrix} I_{p+q} & 0 \\ 0 & 0 \end{pmatrix} \right]^{-1}, \quad \text{with} \quad G_0(z) = G(z), \quad G_1(z) = \widehat{G}(z).$$

Taking the derivative with respect to t, we immediately obtain that

$$\partial_t G_t(z) = zn^{-10} G_t(z) \begin{pmatrix} I_{p+q} & 0 \\ 0 & 0 \end{pmatrix} G_t(z). \tag{5.11}$$

Thus applying Gronwall's inequality to

$$\|G_t(z)\|_{\max} \leqslant \|G(z)\|_{\max} + Cn^{-9} \int_0^t \|G_s(z)\|_{\max}^2 \mathrm{d}s,$$

we get that $\max_{0 \leqslant t \leqslant 1} \|G_t(z)\|_{\max} = \mathrm{O}(1)$ on Ξ. Then using (5.11) again, we get (5.10). □

Note that the bound (5.10) is purely deterministic on Ξ, so we do not lose any probability in this claim. Moreover, such a small error n^{-8} is negligible for our proof.

In the following proof, we use the regularized resolvents with $z = \theta_l + \mathrm{i}n^{-4}$. Then by (5.9), $\widehat{S}^{\{\gamma\}}$, $\widehat{R}^{\{\gamma\}}$ and $\widehat{T}^{\{\gamma\}}$ satisfy the deterministic bounds

$$\max_{\gamma} \max \left\{ \|\widehat{S}^{\{\gamma\}}(z)\|, \|\widehat{T}^{\{\gamma\}}(z)\|, \|\widehat{R}^{\{\gamma\}}(z)\| \right\} \lesssim n^{14}. \tag{5.12}$$

As remarked above, because of this bound, Lemma 3.2 (iii) can be used tacitly, and we will not emphasize this fact again in the proof. Using the expansion (5.8) for a sufficiently large k (for example, $k = 100$ will be enough), $|W_{i\mu}| \prec n^{-1/2}$, the anisotropic local law (3.24) for $\widehat{T}$, and the bound (5.12) for $\widehat{R}$, we can obtain that for any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$,

$$\max_{\gamma} \left| \left\langle \mathbf{u}, \left[\widehat{R}^{\{\gamma\}}(z) - \Pi(z) \right] \mathbf{v} \right\rangle \right| \prec n^{-1/2}. \tag{5.13}$$

Moreover, using the same argument as in the proof of Claim 5.2, we can easily show that

$$\mathcal{M}(\theta_l) \text{ has the same asymptotic distribution as } \widehat{\mathcal{M}}(z), \tag{5.14}$$

where $\widehat{\mathcal{M}}(z)$ is defined as

$$\widehat{\mathcal{M}}(z) := \sqrt{n}\,\mathscr{U}^\top \left[\widehat{G}(z) - \Pi(z) \right] \mathscr{U}, \quad z = \theta_l + \mathrm{i}n^{-4}, \quad \mathscr{U} := \begin{pmatrix} \mathbf{U}_a & 0 & 0 & 0 \\ 0 & \mathbf{U}_b & 0 & 0 \\ 0 & 0 & V & 0 \\ 0 & 0 & 0 & V \end{pmatrix}. \tag{5.15}$$

Moreover, by replacing $\widehat{G}$ with $\widehat{G}^g$ or $\widehat{G}^{\{\gamma\}}$, we can also define $\widehat{\mathcal{M}}^g$ or $\widehat{\mathcal{M}}^{\{\gamma\}}$. Then we will use the following comparison lemma to complete the proof of Theorem 2.3.

Lemma 5.3. *Fix any $\gamma = \Phi(i, \mu)$ with $(i, \mu) \in \mathcal{J}_s$. We abbreviate*

$$\mathcal{M}_R^{\{\gamma\}} := \sqrt{n}\,\mathscr{U}^\top \left[\widehat{R}^{\{\gamma\}}(z) - \Pi(z) \right] \mathscr{U}, \quad z = \theta_l + \mathrm{i}n^{-4}.$$

The matrices $\mathcal{M}_S^{\{\gamma\}}$ and $\mathcal{M}_T^{\{\gamma\}}$ are defined similarly by replacing $\widehat{R}^{\{\gamma\}}$ with $\widehat{S}^{\{\gamma\}}$ and $\widehat{T}^{\{\gamma\}}$, respectively. Let $f \in C_b^3(\mathbb{C}^{4r\times 4r})$ be any function with bounded partial derivatives up to third order, and $a \equiv a_n$ be an arbitrary deterministic sequence of $4r \times 4r$ symmetric matrices. Then we have that

$$\mathbb{E}f\left(\mathcal{M}_T^{\{\gamma\}} + a\right) = \mathbb{E}f\left(\mathcal{M}_R^{\{\gamma\}} + a\right) + \sum_{k,l=1}^{4r} \mathcal{Q}_{kl}^{\{\gamma\}} \mathbb{E}\frac{\partial f}{\partial x_{kl}}\left(\mathcal{M}_R^{\{\gamma\}} + a\right) + \mathscr{A}_\gamma + \mathrm{O}_\prec(n^{-\tau_0}\mathcal{E}_\gamma), \tag{5.16}$$

and

$$\mathbb{E}f\left(\mathcal{M}_S^{\{\gamma\}} + a\right) = \mathbb{E}f\left(\mathcal{M}_R^{\{\gamma\}} + a\right) + \mathscr{A}_\gamma + \mathrm{O}_\prec(n^{-\tau_0}\mathcal{E}_\gamma), \tag{5.17}$$

where $\mathscr{A}_\gamma$ satisfies $\mathscr{A}_\gamma \prec n^{-\tau_0}$, and we denote

$$\mathcal{Q}_{kl}^{\{\gamma\}} := \begin{cases} -n^{-1}\mu_x^{(3)}\left(\mathscr{U}_{\mu k}\mathscr{U}_{il} + \mathscr{U}_{ik}\mathscr{U}_{\mu l}\right), & \text{if } \ \mu \in \mathcal{I}_3 \\ -n^{-1}\mu_y^{(3)}\left(\mathscr{U}_{\mu k}\mathscr{U}_{il} + \mathscr{U}_{ik}\mathscr{U}_{\mu l}\right), & \text{if } \ \mu \in \mathcal{I}_4 \end{cases},$$

and

$$\mathcal{E}_\gamma := \sum_{k,l=1}^{4r} \sum_{\sigma_1,\sigma_2=0}^{2} n^{-2+\sigma_1/2+\sigma_2/2} |\mathscr{U}_{ik}|^{\sigma_1} |\mathscr{U}_{\mu l}|^{\sigma_2}. \tag{5.18}$$

Here recall that $\mu_x^{(3)}$ and $\mu_y^{(3)}$ are defined in (4.32).

Proof. The proof of this lemma is almost the same as the one for Lemma 7.13 of [26], where the main inputs are the local laws (3.24) and (5.13), the simple identity (5.6), and the resolvent expansions (5.7) and (5.8). The cosmetic modifications are mainly due to the fact that our local law takes a different form than the one in Theorem 2.2 of [26]. So we ignore the details. □

Combining Proposition 3.11, Proposition 4.1 and Lemma 5.3, we can conclude the proof of Theorem 2.3.

Proof of Theorem 2.3. We fix any function $f \in C_c^\infty(\mathbb{C}^{4r\times 4r})$ and V satisfying (4.4) and (4.5). Using (5.16) and (5.17), we get that

$$\mathbb{E}_{X,Y} f\left(\mathcal{M}_T^{\{\gamma\}} + a\right) = \mathbb{E}_{X,Y} f\left(\mathcal{M}_S^{\{\gamma\}} + a\right) + \sum_{k,l=1}^{4r} \mathcal{Q}_{kl}^{\{\gamma\}} \mathbb{E}_{X,Y}\frac{\partial f}{\partial x_{kl}}\left(\mathcal{M}_R^{\{\gamma\}} + a\right) + \mathrm{O}_\prec(n^{-\tau_0}\mathcal{E}_\gamma), \tag{5.19}$$

where $\mathbb{E}_{X,Y}$ means the partial expectation with respect to the randomness of X, Y, X^g and Y^g (for simplicity, we did not add X^g and Y^g to the subscript). Since $|\mathscr{U}_{\mu k}| \leqslant n^{-1/2+\varepsilon}$ for $\mu \in \mathcal{I}_3 \cup \mathcal{I}_4$ and $|\mathscr{U}_{il}| \leqslant n^{-\tau_0}$ for $i \in \mathcal{I}_s$, it is easy to check that

$$\|\mathcal{Q}^{\{\gamma\}}\|_{\max} \lesssim \min\{n^{-3/2-\tau_0+\varepsilon}, \mathcal{E}_\gamma\}, \quad \text{for} \quad 1 \leqslant \gamma \leqslant \gamma_{\max},$$

where $\mathcal{Q}^{\{\gamma\}}$ is the $4r \times 4r$ matrix with entries $\mathcal{Q}_{kl}^{\{\gamma\}}$. Thus for any fixed $1 \leqslant k,l \leqslant 4r$ and $1 \leqslant \gamma \leqslant \gamma_{\max}$, applying (5.16) with f replaced by $\partial_{x_{kl}} f$, we get that

$$\mathbb{E}_{X,Y}\frac{\partial f}{\partial x_{kl}}\left(\mathcal{M}_R^{\{\gamma\}} + a\right) = \mathbb{E}_{X,Y}\frac{\partial f}{\partial x_{kl}}\left(\mathcal{M}_T^{\{\gamma\}} + a\right) + \mathrm{O}_\prec(n^{-\tau_0}).$$

Plugging it into (5.19), we get that

$$\mathbb{E}_{X,Y} f\left(\mathcal{M}_S^{\{\gamma\}} + a\right) = \mathbb{E}_{X,Y} f\left(\mathcal{M}_T^{\{\gamma\}} + a\right) - \sum_{k,l=1}^{4r} \mathcal{Q}_{kl}^{\{\gamma\}} \mathbb{E}_{X,Y}\frac{\partial f}{\partial x_{kl}}\left(\mathcal{M}_T^{\{\gamma\}} + a\right) + \mathrm{O}_\prec(n^{-\tau_0}\mathcal{E}_\gamma).$$

On the other hand, we have the Taylor expansion

$$\mathbb{E}_{X,Y} f\left(\mathcal{M}_T^{\{\gamma\}} + a - \mathcal{Q}^{\{\gamma\}}\right) = \mathbb{E}_{X,Y} f\left(\mathcal{M}_T^{\{\gamma\}} + a\right) - \sum_{k,l=1}^{4r} \mathcal{Q}_{kl}^{\{\gamma\}} \mathbb{E}_{X,Y} \frac{\partial f}{\partial x_{kl}} \left(\mathcal{M}_T^{\{\gamma\}} + a\right) + \mathrm{O}_\prec(n^{-\tau_0}\mathcal{E}_\gamma).$$

Comparing the above two equations, we get that

$$\mathbb{E}_{X,Y} f\left(\mathcal{M}_T^{\{\gamma\}} + a - \mathcal{Q}^{\{\gamma\}}\right) = \mathbb{E}_{X,Y} f\left(\mathcal{M}_S^{\{\gamma\}} + a\right) + \mathrm{O}_\prec(n^{-\tau_0}\mathcal{E}_\gamma). \tag{5.20}$$

We iterate (5.20) starting at $\gamma = 1$ and $a = 0$, and obtain that

$$\mathbb{E}_{X,Y} f\left(\mathcal{M}_T^{(\gamma_{\max})} - \sum_{\gamma=1}^{\gamma_{\max}} \mathcal{Q}^{\{\gamma\}}\right) = \mathbb{E}_{X,Y} f\left(\mathcal{M}_T^{(0)}\right) + \mathrm{O}_\prec(n^{-\tau_0}), \tag{5.21}$$

where we also used the bound $\sum_\gamma \mathcal{E}_\gamma = \mathrm{O}(1)$, which can be verified directly from the definition (5.18). Now using (4.34), we can bound that

$$\sum_{\gamma=1}^{\gamma_{\max}} \mathcal{Q}^{\{\gamma\}} \prec n^{-1/2}.$$

Plugging it into (5.21), we obtain that

$$\mathbb{E} f\left(\mathcal{M}_T^{(\gamma_{\max})}\right) = \mathbb{E} f\left(\mathcal{M}_T^{(0)}\right) + \mathrm{O}_\prec(n^{-\tau_0}).$$

This shows that $\widehat{\mathcal{M}}(z)$ has the same asymptotic distribution as $\widehat{\mathcal{M}}^g(z)$ in the almost Gaussian case. Combining this fact with (5.14), Proposition 3.11 and Proposition 4.1, we conclude (2.31) when f is smooth. Extension to any bounded continuous f follows from a standard argument. □

6 Proof of Theorem 2.4

In this section, we present the proof of Theorem 2.4 based on a comparison with the case in Theorem 2.3. We first truncate the entries of X, Y and Z using the moment condition (2.34). Choose a constant $c_\phi > 0$ small enough such that $(n^{1/4-c_\phi})^{8+c_0} \geqslant n^{2+\varepsilon_0}$ and $(n^{1/4-c_\phi})^{4+c_0} \geqslant n^{1+\varepsilon_0}$ for a constant $\varepsilon_0 > 0$. Then we introduce the following truncations on the entries of X, Y and Z:

$$X'_{ij} = \mathbf{1}_{|X_{ij}| \leqslant n^{-1/4-c_\phi}} X_{ij}, \quad Y'_{ij} = \mathbf{1}_{|Y_{ij}| \leqslant n^{-1/4-c_\phi}} Y_{ij}, \quad Z'_{ij} = \mathbf{1}_{|Z_{ij}| \leqslant n^{-1/4-c_\phi}} Z_{ij}.$$

In other words, we restrict ourselves to the following event:

$$\Omega := \left\{ \max_{i,j} |X_{ij}| \leqslant \phi_n, \max_{i,j} |Y_{ij}| \leqslant \phi_n, \max_{i,j} |Z_{ij}| \leqslant \phi_n \right\}, \quad \text{with} \quad \phi_n := n^{-1/4-c_\phi}.$$

Combining the condition (2.34) with Markov's inequality, and using a simple union bound, we get that

$$\mathbb{P}(X' \neq X, Y' \neq Y, Z' \neq Z) = \mathrm{O}(n^{-\varepsilon_0}). \tag{6.1}$$

Using (2.34) and integration by parts, it is easy to verify that

$$\mathbb{E}\,|X_{ij}|\, 1_{|X_{ij}|>\phi_n} = \mathrm{O}(n^{-2-\varepsilon_0}), \quad \mathbb{E}\,|X_{ij}|^2\, 1_{|X_{ij}|>\phi_n} = \mathrm{O}(n^{-2-\varepsilon_0}),$$

which implies that

$$|\mathbb{E}X'_{ij}| = \mathrm{O}(n^{-2-\varepsilon_0}), \quad \mathbb{E}|X'_{ij}|^2 = n^{-1} + \mathrm{O}(n^{-2-\varepsilon_0}). \tag{6.2}$$

Moreover, we trivially have that

$$\mathbb{E}|X'_{ij}|^4 \leqslant \mathbb{E}|X_{ij}|^4 = \mathrm{O}(n^{-2}).$$

Similar estimates also hold for the entries of Y and Z. Then we introduce the matrices

$$\mathring{X} = \frac{X' - \mathbb{E}X'}{\mathrm{Var}(X'_{11})}, \quad \mathring{Y} = \frac{Y' - \mathbb{E}Y'}{\mathrm{Var}(Y'_{11})}, \quad \mathring{Z} = \frac{Z' - \mathbb{E}Z'}{\mathrm{Var}(Z'_{11})}.$$

Note that by (6.2), we have the estimates

$$\|\mathbb{E}X'\| = \mathrm{O}(n^{-1-\varepsilon_0}), \quad \mathrm{Var}(X'_{11}) = n^{-1}\left[1 + \mathrm{O}(n^{-1-\varepsilon_0})\right], \tag{6.3}$$

and similar estimates also hold for $\|\mathbb{E}Y'\|$, $\mathrm{Var}(Y'_{11})$, $\|\mathbb{E}Z'\|$ and $\mathrm{Var}(Z'_{11})$. Now we define the SCC matrices $\mathring{\mathcal{C}}_{\mathcal{X}\mathcal{Y}}$ and $\mathring{\mathcal{C}}_{XY}$ by replacing (X, Y, Z) with $(\mathring{X}, \mathring{Y}, \mathring{Z})$ in (2.10) and (2.11). With the estimate (6.3), we can readily bound the differences between the eigenvalues of $\mathring{\mathcal{C}}_{\mathcal{X}\mathcal{Y}}$ and those of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ using Weyl's inequality.

Lemma 6.1. *Under the above setting, we have that*

$$\mathbb{P}\left(\left\|\mathcal{C}_{\mathcal{X}\mathcal{Y}} - \mathring{\mathcal{C}}_{\mathcal{X}\mathcal{Y}}\right\| = \mathrm{O}\left(n^{-1-\varepsilon_0}\right)\right) = 1 - \mathrm{O}\left(n^{-\varepsilon_0}\right).$$

Proof. This lemma is an easy consequence of (6.3) and the singular value bounds in (6.10) and (6.11) below. Moreover, the probability bound is due to (6.1). □

By the above lemma, it suffices to prove that Theorem 2.4 holds under the following assumptions on (X, Y, Z), which correspond to the above setting for $(\mathring{X}, \mathring{Y}, \mathring{Z})$.

Assumption 6.2. *Assume that $X = (X_{ij})$, $Y = (Y_{ij})$ and $Z = (Z_{ij})$ are independent $p \times n$, $q \times n$ and $r \times n$ matrices, whose entries are real i.i.d. random variables satisfying* (2.1), (2.2), *the bounded fourth moment condition*

$$\max\left\{\mathbb{E}|X_{11}|^4, \mathbb{E}|Y_{11}|^4, \mathbb{E}|Z_{11}|^4\right\} \lesssim n^{-2}, \tag{6.4}$$

and the following bounded support condition with $\phi_n = n^{-1/4-c_\phi}$:

$$\max\left\{\max_{i,j}|X_{ij}|, \max_{i,j}|Y_{ij}|, \max_{i,j}|Z_{ij}|\right\} \leqslant \phi_n. \tag{6.5}$$

Moreover, we assume that Assumption 2.1 (iii)–(iv) hold.

The local laws in Section 3.2 can be extended to the above setting. More precisely, we have proved the following theorem in [31, 40].

Theorem 6.3. *Suppose Assumption 6.2 holds.*

(i) (Outliers locations: Theorem 2.9 of [31]) If $t_i \geqslant t_c + n^{-1/3} + \phi_n$, then we have that

$$|\widetilde{\lambda}_i - \theta_i| \prec n^{-1/2}|t_i - t_c|^{1/2} + \phi_n|t_i - t_c|. \tag{6.6}$$

On the other hand, for any $i = \mathrm{O}(1)$ with $t_i < t_c + n^{-1/3} + \phi_n$, we have that

$$|\widetilde{\lambda}_i - \lambda_+| \prec n^{-2/3} + \phi_n^2. \tag{6.7}$$

(ii) (Anisotropic local law: Theorem 2.11 of [40] and Theorem 3.10 of [31]) For any fixed $\varepsilon > 0$ and deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$, the following estimates hold. For $z \in S(\varepsilon)$, we have that

$$|\langle \mathbf{u}, G(z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec \phi_n + \Psi(z) + (n\eta)^{-1}. \tag{6.8}$$

For $z \in S_{out}(\varepsilon)$, we have that

$$|\langle \mathbf{u}, G(z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec \phi_n + n^{-1/2}(\kappa + \eta)^{-1/4}. \tag{6.9}$$

(iii) (Eigenvalue rigidity: Theorem 2.5 of [40]) The eigenvalue rigidity estimate (3.4) *holds.*

(iv) (Singular value bounds: Lemma 3.3 of [40]) For any constant $\varepsilon > 0$, we have that with high probability,

$$(1 - \sqrt{c_1})^2 - \varepsilon \leqslant \lambda_p(S_{xx}) \leqslant \lambda_1(S_{xx}) \leqslant (1 + \sqrt{c_1})^2 + \varepsilon, \tag{6.10}$$

and

$$(1 - \sqrt{c_2})^2 - \varepsilon \leqslant \lambda_q(S_{yy}) \leqslant \lambda_1(S_{yy}) \leqslant (1 + \sqrt{c_2})^2 + \varepsilon. \tag{6.11}$$

For the above results to hold, it is not necessary to assume that the entries of X, Y and Z are identically distributed, that is, only independence and moment conditions are needed.

Moreover, Lemma 4.3 takes the following general form.

Lemma 6.4 (Lemma 3.8 of [15])**.** *Let (x_i), (y_j) be independent families of centered independent random variables, and $(\mathcal{A}_i)$, $(\mathcal{B}_{ij})$ be families of deterministic complex numbers. Suppose the entries x_i, y_j have variances at most n^{-1} and satisfy the bounded support condition (6.5). Then the following large deviation bounds hold:*

$$\Big|\sum_i \mathcal{A}_i x_i\Big| \prec \phi_n \max_i |\mathcal{A}_i| + \frac{1}{\sqrt{n}}\Big(\sum_i |\mathcal{A}_i|^2\Big)^{1/2}, \quad \Big|\sum_{i,j} x_i \mathcal{B}_{ij} y_j\Big| \prec \phi_n^2 \mathcal{B}_d + \phi_n \mathcal{B}_o + \frac{1}{n}\Big(\sum_{i \neq j} |\mathcal{B}_{ij}|^2\Big)^{1/2},$$

$$\Big|\sum_i \overline{x}_i \mathcal{B}_{ii} x_i - \sum_i (\mathbb{E}|x_i|^2)\mathcal{B}_{ii}\Big| \prec \phi_n \mathcal{B}_d, \quad \Big|\sum_{i \neq j} \overline{x}_i \mathcal{B}_{ij} x_j\Big| \prec \phi_n \mathcal{B}_o + \frac{1}{n}\Big(\sum_{i \neq j} |\mathcal{B}_{ij}|^2\Big)^{1/2},$$

where $\mathcal{B}_d := \max_i |\mathcal{B}_{ii}|$ and $\mathcal{B}_o := \max_{i \neq j} |\mathcal{B}_{ij}|$.

Following the arguments in Section 3.3 and using Theorem 6.3, we can obtain a similar equation as (3.37):

$$\det\big[f_c(\lambda) I_r - \operatorname{diag}(t_1, \cdots, t_r) + \mathcal{O}^\top \mathcal{E}_r(\lambda)\mathcal{O} + \mathrm{O}_\prec(n^{-1} + \phi_n^2)\big] = 0. \tag{6.12}$$

Then using (6.12) and (6.6), as in Proposition 3.11, we can get that

$$\Big|\Big(\widetilde{\lambda}_{\alpha(i)} - \theta_l\Big) - \mu_i\Big\{a(t_l)\big[\operatorname{diag}(t_1, \cdots, t_r) - t_l - \mathcal{O}^\top \mathcal{E}_r(\theta_l)\mathcal{O}\big]_{[\![\gamma(l)]\!]}\Big\}\Big| \prec n^{-1/2-\varepsilon}, \tag{6.13}$$

for a constant $\varepsilon > 0$ depending on c_ϕ only. Again the proof is the same as the one for Proposition 4.5 in [27], so we omit the details. We also remark that this proof is the only place where we need to use the well-separation condition (2.35).

With (6.12), the problem is once again reduced to the study of the CLT of $\mathcal{M}_0(\theta_l)$ in (3.46). Using Lemma 6.4, we can obtain a similar estimate as in (3.32):

$$\|ZZ^\top - I_r\| \prec \phi_n, \tag{6.14}$$

Thus similar to (4.4), we can introduce an $n \times r$ partial orthogonal matrix V such that

$$V^\top V = I_r, \quad \|V - Z^\top\|_F \prec \phi_n. \tag{6.15}$$

With (6.15) and (6.9), we can check that

$$\|\mathcal{M}(\theta_l) - \mathcal{M}_0(\theta_l)\| \prec \sqrt{n}\phi_n^2 \leqslant n^{-2c_\phi},$$

where the matrix $\mathcal{M}$ is defined in (4.7). Thus to prove Theorem 2.4, it suffices to prove the CLT for $\mathcal{M}(\theta_l)$. As in Section 5, to avoid singular behaviours of the resolvents on exceptional low-probability events, we will use the regularized resolvent $\widehat{G}(z)$ in Definition 5.1 with $z = \theta_l + \mathrm{i}n^{-4}$ throughout the rest of the proof. However, for simplicity of notations, we still use the notation $G(z)$ to denote the regularized resolvents in the following proof, while keeping in mind that the bound (5.12) will hold for all the resolvent entries appearing below and hence Lemma 3.2 (iii) can be applied without worry. Finally, we remark that the rest of the proof will be conditional on Z and V, i.e. they are regarded as deterministic matrices unless specified otherwise.

Given any random matrices X and Y satisfying Assumption 6.2, we can construct matrices $\widetilde{X}$ and $\widetilde{Y}$, whose entries match the first four moments of the entries of X and Y but with smaller support $n^{-1/2}$.

Lemma 6.5 (Lemma 5.1 of [29]). *Suppose X, Y and Z satisfy Assumption 6.2. Then there exist independent random matrices $\widetilde{X} = (\widetilde{X}_{ij})$, $\widetilde{Y} = (\widetilde{Y}_{ij})$ and $\widetilde{Z} = (\widetilde{Z}_{ij})$ satisfying Assumption 6.2, such that the condition (6.5) holds with ϕ_n replaced by $n^{-1/2}$. Moreover, they satisfy the following moment matching conditions:*

$$\mathbb{E}X_{ij}^k = \mathbb{E}\widetilde{X}_{ij}^k, \quad \mathbb{E}Y_{ij}^k = \mathbb{E}\widetilde{Y}_{ij}^k, \quad \mathbb{E}Z_{ij}^k = \mathbb{E}\widetilde{Z}_{ij}^k, \quad k = 1, 2, 3, 4. \tag{6.16}$$

Note that $\widetilde{X}$, $\widetilde{Y}$ and $\widetilde{Z}$ satisfy the setting of Theorem 2.3. By replacing (X, Y) with $(\widetilde{X}, \widetilde{Y})$ in (3.6), (3.7) and (4.7), We can define $\widetilde{H}(z)$, $\widetilde{G}(z)$ and $\widetilde{\mathcal{M}}(z)$. In Section 5, we have proved the CLT for $\widetilde{\mathcal{M}}(\theta_l)$. The rest of the proof is devoted to showing that $\mathcal{M}(\theta_l)$ has the same asymptotic distribution as $\widetilde{\mathcal{M}}(\theta_l)$.

Proposition 6.6. *Suppose Assumption 6.2 holds. Let $\widetilde{X}$ and $\widetilde{Y}$ be two random matrices constructed as in Lemma 6.5. Then there exists a constant $\varepsilon > 0$ such that for any function $f \in C_c^\infty(\mathbb{C}^{4r\times 4r})$, we have*

$$\mathbb{E}f\left(\mathcal{M}(z)\right) = \mathbb{E}f(\widetilde{\mathcal{M}}(z)) + \mathrm{O}(n^{-\varepsilon}), \quad \textit{for} \quad z = \theta_l + \mathrm{i}n^{-4}.$$

To prove this proposition, we will use the continuous comparison method introduced in [28]. We first introduce the following interpolation between (X, Y) and $(\widetilde{X}, \widetilde{Y})$.

Definition 6.7 (Interpolating matrices). *Introduce the notations $X^0 := \widetilde{X}$ and $X^1 := X$. Let $\rho_{i\mu}^0$ and $\rho_{i\mu}^1$ be the laws of $\widetilde{X}_{i\mu}$ and $X_{i\mu}$, respectively. For $\theta \in [0,1]$, we define the interpolated law*

$$\rho_{i\mu}^\theta := (1-\theta)\rho_{i\mu}^0 + \theta\rho_{i\mu}^1.$$

Let $\{X^\theta : \theta \in (0,1)\}$ be a collection of random matrices such that the following properties hold. For any fixed $\theta \in (0,1)$, (X^0, X^θ, X^1) is a triple of independent $\mathcal{I}_1 \times \mathcal{I}_3$ random matrices, and the matrix $X^\theta = (X_{i\mu}^\theta)$ has law

$$\prod_{i\in\mathcal{I}_1}\prod_{\mu\in\mathcal{I}_3} \rho_{i\mu}^\theta\left(\mathrm{d}X_{i\mu}^\theta\right). \tag{6.17}$$

Note that we do not require X^{θ_1} to be independent of X^{θ_2} for $\theta_1 \neq \theta_2 \in (0,1)$. For $\lambda \in \mathbb{R}$, $i \in \mathcal{I}_1$ and $\mu \in \mathcal{I}_3$, we define the matrix $X_{(i\mu)}^{\theta,\lambda}$ through

$$\left(X_{(i\mu)}^{\theta,\lambda}\right)_{j\nu} := \begin{cases} X_{i\mu}^\theta, & \textit{if } (j,\nu) \neq (i,\mu) \\ \lambda, & \textit{if } (j,\nu) = (i,\mu) \end{cases}. \tag{6.18}$$

In a similar way, we can define the collection of random matrices $\{Y^\theta : \theta \in [0,1]\}$ for all $\theta \in [0,1]$ with $Y^0 := \widetilde{Y}$ and $Y^1 := Y$. We require that for any fixed $\theta \in (0,1)$, Y^θ is independent of $(X^0, X^\theta, X^1, Y^0, Y^1)$. For $\lambda \in \mathbb{R}$, $i \in \mathcal{I}_2$ and $\mu \in \mathcal{I}_4$, we define $Y^{\theta,\lambda}_{(i\mu)}$ in the same way as (6.18). We also introduce the resolvents

$$G^\theta(z) := G\left(X^\theta, Y^\theta, z\right), \quad G^{\theta,\lambda}_{(i\mu)}(z) := \begin{cases} G\left(X^{\theta,\lambda}_{(i\mu)}, Y^\theta, z\right), & \text{if } i \in \mathcal{I}_1, \mu \in \mathcal{I}_3 \\ G\left(X^\theta, Y^{\theta,\lambda}_{(i\mu)}, z\right), & \text{if } i \in \mathcal{I}_2, \mu \in \mathcal{I}_4 \end{cases}.$$

Using (6.17) and fundamental calculus, it is easy to derive the following basic interpolation formula.

Lemma 6.8. *For any differentiable function $F : \mathbb{C}^{\mathcal{I}_1 \times \mathcal{I}_3} \times \mathbb{C}^{\mathcal{I}_2 \times \mathcal{I}_4} \to \mathbb{C}$, we have that*

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\theta} \mathbb{E} F(X^\theta, Y^\theta) = &\sum_{i \in \mathcal{I}_1, \mu \in \mathcal{I}_3} \left[\mathbb{E} F\left(X^{\theta, X^1_{i\mu}}_{(i\mu)}, Y^\theta\right) - \mathbb{E} F\left(X^{\theta, X^0_{i\mu}}_{(i\mu)}, Y^\theta\right) \right] \\ &+ \sum_{i \in \mathcal{I}_2, \mu \in \mathcal{I}_4} \left[\mathbb{E} F\left(X^\theta, Y^{\theta, Y^1_{i\mu}}_{(i\mu)}\right) - \mathbb{E} F\left(X^\theta, Y^{\theta, Y^0_{i\mu}}_{(i\mu)}\right) \right], \end{aligned} \tag{6.19}$$

provided all the expectations exist.

We shall apply Lemma 6.8 to $F(X^\theta, Y^\theta) = f(\mathcal{M}\left(X^\theta, Y^\theta, z\right))$ for the function f in Proposition 6.6, where $\mathcal{M}\left(X^\theta, Y^\theta, z\right)$ is defined by replacing $G(z) \equiv G(X, Y, z)$ with $G^\theta(z) \equiv G(X^\theta, Y^\theta, z)$. The main work is to show the following estimate for the right-hand side of (6.19).

Lemma 6.9. *Under the assumptions of Proposition 6.6, there exists a constant $\varepsilon > 0$ such that*

$$\sum_{i \in \mathcal{I}_1} \sum_{\mu \in \mathcal{I}_3} \left[\mathbb{E} f\left(\mathcal{M}\left(X^{\theta, X^1_{i\mu}}_{(i\mu)}, Y^\theta\right)\right) - \mathbb{E} f\left(\mathcal{M}\left(X^{\theta, X^0_{i\mu}}_{(i\mu)}, Y^\theta\right)\right) \right] = \mathrm{O}(n^{-\varepsilon}), \tag{6.20}$$

and

$$\sum_{i \in \mathcal{I}_2} \sum_{\mu \in \mathcal{I}_4} \left[\mathbb{E} f\left(\mathcal{M}\left(X^\theta, Y^{\theta, Y^1_{i\mu}}_{(i\mu)}\right)\right) - \mathbb{E} f\left(\mathcal{M}\left(X^\theta, Y^{\theta, Y^0_{i\mu}}_{(i\mu)}\right)\right) \right] = \mathrm{O}(n^{-\varepsilon}), \tag{6.21}$$

for all $\theta \in [0,1]$.

Combining Lemma 6.8 and Lemma 6.9, we conclude Proposition 6.6. The proof of Lemma 6.9 is based on an expansion approach. As in (5.7) and (5.8), for any $i \in \mathcal{I}_1$, $\mu \in \mathcal{I}_3$, $\lambda, \lambda' \in \mathbb{R}$ and $K \in \mathbb{N}$, we have the resolvent expansion

$$G^{\theta,\lambda'}_{(i\mu)} = G^{\theta,\lambda}_{(i\mu)} + \sum_{k=1}^{K} (\lambda - \lambda')^k G^{\theta,\lambda}_{(i\mu)} \left(E^{\{i,\mu\}} G^{\theta,\lambda}_{(i\mu)}\right)^k + (\lambda - \lambda')^{K+1} G^{\theta,\lambda'}_{(i\mu)} \left(E^{\{i,\mu\}} G^{\theta,\lambda}_{(i\mu)}\right)^{K+1}, \tag{6.22}$$

where $E^{\{i,\mu\}}$ is the matrix defined by $(E^{\{i,\mu\}})_{ab} = \mathbf{1}_{(a,b)=(i,\mu)} + \mathbf{1}_{(a,b)=(\mu,i)}$ as in (5.4). With this expansion, we can readily obtain the following estimate: if y is a random variable satisfying $|y| \leqslant \phi_n$, then for any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$, we have that

$$\left\langle \mathbf{u}, \left[G^{\theta,y}_{(i\mu)}(z) - \Pi(z)\right] \mathbf{v} \right\rangle \prec \phi_n, \quad z = \theta_l + \mathrm{i} n^{-4}. \tag{6.23}$$

In fact, to prove this estimate, we will apply the expansion (6.22) for a sufficiently large K, say $K = 100$, with $\lambda' = y$ and $\lambda = X^\theta_{i\mu}$, so that $G^{\theta,\lambda}_{(i\mu)} = G^\theta$. Then to bound the resulting expansion on the right-hand side of (6.22), we will use $y \leqslant \phi_n$, $|X^\theta_{i\mu}| \leqslant \phi_n$, the anisotropic local law (6.9) for G^θ, and the rough bound as in (5.12) for $G^{\theta,y}_{(i\mu)}$ in the last term.

Proof Lemma 6.9. We only give the proof of (6.20), while (6.21) obviously can be proved in the same way. For simplicity of notations, we only provide the proof for a simpler version of (6.20),

$$\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_3}\left[\mathbb{E}f\left(M\left(X_{(i\mu)}^{\theta,X_{i\mu}^1},Y^\theta\right)\right)-\mathbb{E}f\left(M\left(X_{(i\mu)}^{\theta,X_{i\mu}^0},Y^\theta\right)\right)\right]=\mathrm{O}(n^{-\varepsilon}),\tag{6.24}$$

where M is defined as

$$M(X,Y):=\sqrt{n}\langle\mathbf{u},(G(X,Y,z)-\Pi(z))\,\mathbf{v}\rangle$$

for some deterministic unit vectors $\mathbf{u},\mathbf{v}\in\mathbb{C}^{\mathcal{I}}$ satisfying that

$$\max_{\mu\in\mathcal{I}_3\cup\mathcal{I}_4}|u(\mu)|\prec\phi_n,\qquad\max_{\mu\in\mathcal{I}_3\cup\mathcal{I}_4}|v(\mu)|\prec\phi_n.\tag{6.25}$$

The proof for (6.20) is exactly the same, except that we need to use multivariable Taylor expansions. Here the condition (6.25) is due to the corresponding bound on V,

$$\|V\|_{\max}\leqslant\|V-Z\|_{\max}+\|Z\|_{\max}\prec\phi_n$$

by (6.15) and the bounded support condition in (6.5).

In the following proof, for simplicity of notations, we fix a $\theta\in[0,1]$ and denote $M_{(i\mu)}(\lambda):=M(X_{(i\mu)}^{\theta,\lambda})$ while ignoring Y^θ from the argument. Recall that $\phi_n=n^{-1/4-c_\phi}$. Using (6.22) with $K=9$ and the local law (6.23), we get that for a random variable y satisfying $|y|\leqslant\phi_n$,

$$M_{(i\mu)}(y)-M_{(i\mu)}(0)=\sum_{k=1}^{9}n^{1/2}(-y)^k x_k(i,\mu)+\mathrm{O}_\prec(n^{-2-10c_\phi}),\tag{6.26}$$

where

$$x_k(i,\mu):=\left\langle\mathbf{u},G_{(i\mu)}^{\theta,0}\left(E^{\{i,\mu\}}G_{(i\mu)}^{\theta,0}\right)^k\mathbf{v}\right\rangle.$$

By (6.23), we have $x_k(i,\mu)\prec 1$ for $k\geqslant 1$. On the other hand, for $k=1$, using (6.23) and (6.25) we can get a better bound

$$x_1(i,\mu)=\left\langle\mathbf{u},G_{(i\mu)}^{\theta,0}E^{\{i,\mu\}}G_{(i\mu)}^{\theta,0}\,\mathbf{v}\right\rangle=\left\langle\mathbf{u},\Pi E^{\{i,\mu\}}\Pi\,\mathbf{v}\right\rangle+\mathrm{O}_\prec(\phi_n)\prec\phi_n.\tag{6.27}$$

Combining this bound with $|y|\leqslant\phi_n$, from (6.26) we immediately obtain the rough bound

$$M_{(i\mu)}(y)-M_{(i\mu)}(0)\prec n^{1/2}\phi_n^2\leqslant n^{-2c_\phi}.\tag{6.28}$$

Now fix an integer $K\geqslant 1/c_\phi$. Using (6.26) and (6.28), the Taylor expansion of f up to the K-th order gives that for $\alpha\in\{0,1\}$,

$$\begin{aligned}&\mathbb{E}f\left(M_{(i\mu)}(X_{i\mu}^\alpha)\right)-\mathbb{E}f\left(M_{(i\mu)}(0)\right)\\&=\sum_{k=1}^{K}\mathbb{E}\frac{f^{(k)}\left(M_{(i\mu)}(0)\right)}{k!}\left[\sum_{l=1}^{9}n^{1/2}(-X_{i\mu}^\alpha)^l x_l(i,\mu)\right]^k+\mathrm{O}_\prec\left(n^{-2-2c_\phi}\right)\\&=\sum_{k=1}^{K}\sum_{s=1}^{K+2k}\sum_{\mathbf{s}}^{*}n^{k/2}\mathbb{E}(-X_{i\mu}^\alpha)^s\mathbb{E}\frac{f^{(k)}\left(M_{(i\mu)}(0)\right)}{k!}\prod_{l=1}^{k}x_{s_l}(i,\mu)+\mathrm{O}_\prec\left(n^{-2-2c_\phi}\right),\end{aligned}$$

where $\sum_{\mathbf{s}}^*$ means the sum over $\mathbf{s}=(s_1,\cdots,s_k)\in\mathbb{N}^k$ satisfying

$$1\leqslant s_i\leqslant 9,\qquad\sum_{l=1}^{k}l\cdot s_l=s.\tag{6.29}$$

Here for the terms with $s > K + 2k$, we have $n^{k/2}\mathbb{E}(-X^{\alpha}_{i\mu})^s \leqslant n^{-2-2c_\phi}$, so they are included into the errors. Now using the moment matching condition (6.16), we get that

$$\left|\mathbb{E} f\left(M_{(i\mu)}(X^1_{i\mu})\right) - \mathbb{E} f\left(M_{(i\mu)}(X^0_{i\mu})\right)\right| \prec \sum_{k=1}^{K}\sum_{s=5}^{K+2k}\sum_{\mathbf{s}}^{*} n^{k/2-2}\phi_n^{s-4}\mathbb{E}\Big|\prod_{l=1}^{k} x_{s_l}(i,\mu)\Big| + n^{-2-2c_\phi},$$

where we used that $\mathbb{E}|X^{\alpha}_{i\mu}|^s \leqslant \phi_n^{s-4}\mathbb{E}|X^{\alpha}_{i\mu}|^4 \lesssim \phi_n^{s-4} n^{-2}$ for $s \geqslant 5$. Thus to show (6.24), we only need to prove that for any fixed $s \geqslant 5$ and $\mathbf{s} \in \mathbb{N}^k$ satisfying (6.29),

$$\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_3} n^{k/2-2}\phi_n^{s-4}\mathbb{E}\Big|\prod_{l=1}^{k} x_{s_l}(i,\mu)\Big| \prec n^{-\varepsilon} \tag{6.30}$$

for some constant $\varepsilon > 0$. For the proof of (6.30), we will consider three different cases. To ease the notation, we introduce the following notion of generalized entries.

Definition 6.10 (Generalized entries). *For $\mathbf{v}, \mathbf{w} \in \mathbb{C}^{\mathcal{I}}$, $\mathfrak{a} \in \mathcal{I}$ and an $\mathcal{I}\times\mathcal{I}$ matrix $\mathcal{A}$, we shall denote*

$$\mathcal{A}_{\mathbf{v}\mathbf{w}} := \langle \mathbf{v}, \mathcal{A}\mathbf{w}\rangle, \quad \mathcal{A}_{\mathbf{v}\mathfrak{a}} := \langle \mathbf{v}, \mathcal{A}\mathbf{e}_{\mathfrak{a}}\rangle, \quad \mathcal{A}_{\mathfrak{a}\mathbf{w}} := \langle \mathbf{e}_{\mathfrak{a}}, \mathcal{A}\mathbf{w}\rangle, \tag{6.31}$$

where $\mathbf{e}_{\mathfrak{a}}$ is the standard unit vector along the $\mathfrak{a}$-th coordinate axis.

Case 1: Suppose $s_l \geqslant 2$ for all $l = 1, \cdots, k$. Then we have $s \geqslant \max\{2k, 5\}$ and

$$n^{k/2-2}\phi_n^{s-4} = n^{-2+k/2-(s-4)/4} n^{-(s-4)c_\phi} \leqslant n^{-1-c_\phi}. \tag{6.32}$$

On the other hand, using (6.22) with $K = 0$ and (6.23), we get that

$$\left|\left\langle \mathbf{e}_i, G^{\theta,0}_{(i\mu)}\,\mathbf{u}\right\rangle\right| \leqslant \left|G^{\theta}_{i\,\mathbf{u}}\right| + \left|X^{\theta}_{i\mu}\right|\left(\left|\left\langle \mathbf{e}_i, G^{\theta,0}_{(i\mu)}\mathbf{e}_i\right\rangle\right|\left|G^{\theta}_{\mu\,\mathbf{u}}\right| + \left|\left\langle \mathbf{e}_i, G^{\theta,0}_{(i\mu)}\mathbf{e}_\mu\right\rangle\right|\left|G^{\theta}_{i\,\mathbf{u}}\right|\right) \prec \left|G^{\theta}_{i\,\mathbf{u}}\right| + \phi_n\left|G^{\theta}_{\mu\,\mathbf{u}}\right|. \tag{6.33}$$

Similarly, we have that

$$\left|\left\langle \mathbf{e}_\mu, G^{\theta,0}_{(i\mu)}\,\mathbf{u}\right\rangle\right| \prec \left|G^{\theta}_{\mu\,\mathbf{u}}\right| + \phi_n\left|G^{\theta}_{i\,\mathbf{u}}\right|. \tag{6.34}$$

Inserting (6.33) and (6.34) into the definition of $x_l(i,\mu)$, we immediately get that

$$|x_l(i,\mu)| \prec \left|G^{\theta}_{i\,\mathbf{u}}\right|^2 + \left|G^{\theta}_{i\,\mathbf{v}}\right|^2 + \left|G^{\theta}_{\mu\,\mathbf{u}}\right|^2 + \left|G^{\theta}_{\mu\,\mathbf{v}}\right|^2, \quad l \geqslant 1. \tag{6.35}$$

We claim that for any deterministic unit vector $\mathbf{u} \in \mathbb{C}^{\mathcal{I}}$,

$$\sum_{i\in\mathcal{I}_1}\left|G^{\theta}_{i\,\mathbf{u}}\right|^2 \prec 1, \quad \sum_{\mu\in\mathcal{I}_3}\left|G^{\theta}_{\mu\,\mathbf{u}}\right|^2 \prec 1. \tag{6.36}$$

We postpone its proof until we complete the proof of Lemma 6.9. Combining (6.32), (6.35) and (6.36), we can bound that

$$\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_3} n^{k/2-2}\phi_n^{s-4}\mathbb{E}\Big|\prod_{l=1}^{k} x_{s_l}(i,\mu)\Big| \prec \sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_3} n^{-1-c_\phi}\left(\left|G^{\theta}_{i\,\mathbf{u}}\right|^2 + \left|G^{\theta}_{i\,\mathbf{v}}\right|^2 + \left|G^{\theta}_{\mu\,\mathbf{u}}\right|^2 + \left|G^{\theta}_{\mu\,\mathbf{v}}\right|^2\right) \prec n^{-c_\phi}.$$

Case 2: Suppose there are at least two l's such that $s_l = 1$. Without loss of generality, we assume that $s_1 = s_2 = \cdots = s_j = 1$ for some $2 \leqslant j \leqslant k$. Then we have $s \geqslant \max\{2k - j, 5\}$, which gives that

$$n^{k/2-2}\phi_n^{s-4}\mathbb{E}\Big|\prod_{l=1}^{k} x_{s_l}(i,\mu)\Big| \prec n^{k/2-2}\phi_n^{s-4}\phi_n^{j-2}|x_1(i,\mu)|^2 \leqslant n^{-1/2-c_\phi}|x_1(i,\mu)|^2. \tag{6.37}$$

where in the second step we used

$$n^{k/2-2}\phi_n^{s+j-6} = n^{-2+k/2-(s+j-6)/4} n^{-(s+j-6)c_\phi} \leqslant n^{-1/2-c_\phi}.$$

Applying (6.33) and (6.34) to (6.27), we can bound that

$$\begin{aligned} |x_1(i,\mu)| &\prec \left(|G^\theta_{i\mathbf{u}}| + \phi_n |G^\theta_{\mu\mathbf{u}}|\right)\left(|G^\theta_{\mu\mathbf{v}}| + \phi_n |G^\theta_{i\mathbf{v}}|\right) + \left(|G^\theta_{\mu\mathbf{u}}| + \phi_n |G^\theta_{i\mathbf{u}}|\right)\left(|G^\theta_{i\mathbf{v}}| + \phi_n |G^\theta_{\mu\mathbf{v}}|\right) \\ &\lesssim |G^\theta_{i\mathbf{u}}||G^\theta_{\mu\mathbf{v}}| + |G^\theta_{\mu\mathbf{u}}||G^\theta_{i\mathbf{v}}| + \phi_n\left(|G^\theta_{i\mathbf{u}}|^2 + |G^\theta_{i\mathbf{v}}|^2 + |G^\theta_{\mu\mathbf{u}}|^2 + |G^\theta_{\mu\mathbf{v}}|^2\right). \end{aligned} \tag{6.38}$$

Now using (6.36) and (6.38), we get that

$$\begin{aligned} \sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_3} |x_1(i,\mu)|^2 &\prec \sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_3}\left[|G^\theta_{i\mathbf{u}}|^2|G^\theta_{\mu\mathbf{v}}|^2 + |G^\theta_{\mu\mathbf{u}}|^2|G^\theta_{i\mathbf{v}}|^2 + \phi_n^2\left(|G^\theta_{i\mathbf{u}}|^4 + |G^\theta_{i\mathbf{v}}|^4 + |G^\theta_{\mu\mathbf{u}}|^4 + |G^\theta_{\mu\mathbf{v}}|^4\right)\right] \\ &\prec 1 + n\phi_n^2. \end{aligned}$$

Combining this bound with (6.37), we get that

$$\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_3} n^{k/2-2}\phi_n^{s-4}\mathbb{E}\Big|\prod_{l=1}^k x_{s_l}(i,\mu)\Big| \prec n^{-1/2-c_\phi}\cdot n\phi_n^2 \leqslant n^{-3c_\phi}.$$

Case 3: Finally, suppose there is only one l such that $s_l = 1$. Without loss of generality, we assume that $s_1 = 1$ and $s_l \geqslant 2$ for $l = 2, \cdots, k$. Thus we have $s \geqslant \max\{2k-1, 5\}$, which gives that

$$\begin{aligned} n^{k/2-2}\phi_n^{s-4}\mathbb{E}\Big|\prod_{l=1}^k x_{s_l}(i,\mu)\Big| &\prec n^{k/2-2}\phi_n^{s-4}|x_1(i,\mu)|\left(|G^\theta_{i\mathbf{u}}|^2 + |G^\theta_{i\mathbf{v}}|^2 + |G^\theta_{\mu\mathbf{u}}|^2 + |G^\theta_{\mu\mathbf{v}}|^2\right) \\ &\leqslant n^{-3/4-c_\phi}\left(|G^\theta_{i\mathbf{u}}||G^\theta_{\mu\mathbf{v}}| + |G^\theta_{\mu\mathbf{u}}||G^\theta_{i\mathbf{v}}|\right)\left(|G^\theta_{i\mathbf{u}}|^2 + |G^\theta_{i\mathbf{v}}|^2 + |G^\theta_{\mu\mathbf{u}}|^2 + |G^\theta_{\mu\mathbf{v}}|^2\right) \\ &\quad + n^{-3/4-c_\phi}\phi_n\left(|G^\theta_{i\mathbf{u}}|^4 + |G^\theta_{i\mathbf{v}}|^4 + |G^\theta_{\mu\mathbf{u}}|^4 + |G^\theta_{\mu\mathbf{v}}|^4\right), \end{aligned} \tag{6.39}$$

where in the first step we used (6.35), and in the second step we used (6.38) and

$$n^{k/2-2}\phi_n^{s-4} = n^{-2+k/2-(s-4)/4} n^{-(s-4)c_\phi} \leqslant n^{-3/4-c_\phi}.$$

Applying (6.36) and Cauchy-Schwarz inequality to (6.39), we get that

$$\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_3} n^{k/2-2}\phi_n^{s-4}\mathbb{E}\Big|\prod_{l=1}^k x_{s_l}(i,\mu)\Big| \prec n^{-2c_\phi}.$$

Combining the above three cases, we conclude (6.30) with $\varepsilon = c_\phi$, which further implies (6.24). With similar arguments, we can conclude (6.20) and (6.21). □

Proof of (6.36). (6.36) is a simple corollary of the spectral decomposition of the resolvent. First, recalling the notations in (2.12), we define

$$\mathcal{H} := S_{xx}^{-1/2} S_{xy} S_{yy}^{-1/2}, \tag{6.40}$$

and the resolvent

$$R(z) := \begin{pmatrix} R_1 & -z^{-1/2}R_1\mathcal{H} \\ -z^{-1/2}\mathcal{H}^\top R_1 & R_2 \end{pmatrix},$$

where the two blocks R_1 and R_2 are defined as

$$R_1(z) := (\mathcal{C}_{XY} - z)^{-1} = \left(\mathcal{H}\mathcal{H}^\top - z\right)^{-1}, \quad R_2(z) := (\mathcal{C}_{YX} - z)^{-1} = \left(\mathcal{H}^\top\mathcal{H} - z\right)^{-1}. \tag{6.41}$$

Consider a singular value decomposition of $\mathcal{H}$:

$$\mathcal{H} = \sum_{k=1}^{q} \sqrt{\lambda_k}\xi_k\zeta_k^\top, \tag{6.42}$$

where λ_k's are the eigenvalues of the null SCC matrix $\mathcal{C}_{XY}$, and ξ_k's and ζ_k's are respectively the left and right singular vectors. Then the singular value decomposition $R(z)$ is given by

$$R(z) = \sum_{k=1}^{q} \frac{1}{\lambda_k - z} \begin{pmatrix} \xi_k\xi_k^\top & -z^{-1/2}\sqrt{\lambda_k}\xi_k\zeta_k^\top \\ -z^{-1/2}\sqrt{\lambda_k}\zeta_k\xi_k^\top & \zeta_k\zeta_k^\top \end{pmatrix} - \frac{1}{z}\begin{pmatrix} \sum_{k=q+1}^{p}\xi_k\xi_k^\top & 0 \\ 0 & 0 \end{pmatrix}. \tag{6.43}$$

Next, we denote the $(\mathcal{I}_1 \cup \mathcal{I}_2) \times (\mathcal{I}_1 \cup \mathcal{I}_2)$ block of $G(z)$ by $\mathcal{G}_L(z)$, the $(\mathcal{I}_1 \cup \mathcal{I}_2) \times (\mathcal{I}_3 \cup \mathcal{I}_4)$ block by $\mathcal{G}_{LR}(z)$, the $(\mathcal{I}_3 \cup \mathcal{I}_4) \times (\mathcal{I}_1 \cup \mathcal{I}_2)$ block by $\mathcal{G}_{RL}(z)$, and the $(\mathcal{I}_3 \cup \mathcal{I}_4) \times (\mathcal{I}_3 \cup \mathcal{I}_4)$ block by $\mathcal{G}_R(z)$. Using Schur complement formula, we can check that

$$\mathcal{G}_L = \begin{pmatrix} S_{xx}^{-1/2} & 0 \\ 0 & S_{yy}^{-1/2} \end{pmatrix} R(z) \begin{pmatrix} S_{xx}^{-1/2} & 0 \\ 0 & S_{yy}^{-1/2} \end{pmatrix}, \tag{6.44}$$

$$\mathcal{G}_R = \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix} + \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}\begin{pmatrix} X^\top & 0 \\ 0 & Y^\top \end{pmatrix}\mathcal{G}_L\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}\begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}, \tag{6.45}$$

$$\mathcal{G}_{LR}(z) = -\mathcal{G}_L(z)\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}\begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}, \quad \mathcal{G}_{RL}(z) = -\begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}\begin{pmatrix} X^\top & 0 \\ 0 & Y^\top \end{pmatrix}\mathcal{G}_L(z). \tag{6.46}$$

Using the rigidity estimate (3.4) given in Theorem 6.3 (iii), we get that

$$\min_{1 \leqslant k \leqslant p} |\lambda_k - z| \gtrsim 1, \quad z = \theta_l + \mathrm{i}n^{-4}. \tag{6.47}$$

Combining it with the SVD (6.43), we see that $\|R(z)\| = \mathrm{O}(1)$ with high probability. Then using (6.44)–(6.46) and (6.10)–(6.11), we obtain that $\|G(z)\| = \mathrm{O}(1)$ with high probability. Thus we have that for any unit vector $\mathbf{u} \in \mathbb{C}^{\mathcal{I}}$,

$$\sum_{\mathfrak{a}\in\mathcal{I}} |G_{\mathfrak{a}\mathbf{u}}|^2 \leqslant \|GG^*\| = \mathrm{O}(1) \quad \text{with high probability}, \tag{6.48}$$

where G^* denotes the conjugate transpose of G. If $G \equiv \widehat{G}$ is the regularized resolvent, then we can apply Claim 5.2 to get that

$$\sum_{\mathfrak{a}\in\mathcal{I}} \left|\widehat{G}_{\mathfrak{a}\mathbf{u}}\right|^2 = \mathrm{O}(1) \quad \text{with high probability}.$$

The above argument obviously also works for the resolvent G^θ, which concludes (6.36). □

Finally, we can complete the proof of Theorem 2.4 using Proposition 6.6.

Proof of Theorem 2.4. First, suppose X, Y and Z satisfy Assumption 6.2, and let $\widetilde{X}$, $\widetilde{Y}$ and $\widetilde{Z}$ be random matrices constructed in Lemma 6.5. Then Theorem 2.4 holds for the SCC matrix defined with $(\widetilde{X}, \widetilde{Y}, \widetilde{Z})$, because they satisfy the assumptions of Theorem 2.3. By Proposition 6.6 (recall that it is proved for the regularized resolvents following the convention stated above Lemma 6.5), we have that asymptotically

$$\widehat{\mathcal{M}}(z) \stackrel{d}{\sim} \widehat{\widetilde{\mathcal{M}}}(z).$$

By the argument in the proof of Claim 5.2, this implies that $\mathcal{M}(\theta_l)$ and $\widetilde{\mathcal{M}}(\theta_l)$ also have the same asymptotic distribution. Moreover, by classical CLT the asymptotic distribution of $\sqrt{n}\left(ZZ^\top - I_r\right)$ is still given by (3.45), which only depends on the first four moments of the entries of Z. Hence by (6.13) we can conclude Theorem 2.4 for the SCC matrix defined with (X, Y, Z) satisfying Assumption 6.2. Finally, using the cut-off argument at the beginning of this section and Lemma 6.1, we conclude Theorem 2.4. $\square$

A Proof of Lemma 4.5

In this section, we give the proof of Lemma 4.5, which is one key result in the proof of Theorem 2.3. Under the setting of Lemma 4.5, we need to study the CLT of the following matrix:

$$\mathcal{Q}_0 := \sqrt{n}\mathscr{V}_0^\top \left(G^{(\mathbb{T})} - \Pi^{(\mathbb{T})}\right)\mathscr{V}_0, \quad \mathscr{V}_0 \equiv \begin{pmatrix} 0 & 0 & \mathbf{V}_1 & 0 \\ 0 & 0 & 0 & \mathbf{V}_2 \\ \mathbf{V}_3 & 0 & 0 & 0 \\ 0 & \mathbf{V}_4 & 0 & 0 \end{pmatrix} := F\,\mathbf{O}\,.$$

It is easy to check that the matrices $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$ and $\mathbf{V}_4$ are respectively $(p-\widetilde{r})\times r$, $(q-\widetilde{r})\times r$, $(n-r)\times r$ and $(n-r)\times r$ random matrices independent of $G^{(\mathbb{T})}$, and satisfy that with high probability,

$$\mathbf{V}_1^\top \mathbf{V}_1 = c_1 I_r + \mathrm{O}_\prec(n^{-1/2}), \quad \mathbf{V}_2^\top \mathbf{V}_2 = c_2 I_r + \mathrm{O}_\prec(n^{-1/2}), \tag{A.1}$$

$$\mathbf{V}_3^\top \mathbf{V}_3 = I_r + \mathrm{O}_\prec(n^{-1/2+2\tau_0}), \quad \mathbf{V}_4^\top \mathbf{V}_4 = I_r + \mathrm{O}_\prec(n^{-1/2+2\tau_0}), \quad \mathbf{V}_3^\top \mathbf{V}_4 = \mathrm{O}_\prec(n^{-1/2+2\tau_0}). \tag{A.2}$$

These conditions all follow from (4.17) and (4.10). For simplicity of notations, we permute the columns of $\mathscr{V}_0$ and study the CLT of

$$\begin{pmatrix} 0 & I_{2r} \\ I_{2r} & 0 \end{pmatrix} \sqrt{n}\mathscr{V}_0^\top (G^{(\mathbb{T})} - \Pi)\mathscr{V}_0 \begin{pmatrix} 0 & I_{2r} \\ I_{2r} & 0 \end{pmatrix} \tag{A.3}$$

Moreover, with a slight abuse of notation, we rename the Gaussian matrices $(X^{(\mathbb{T})}, Y^{(\mathbb{T})})$ in $G^{(\mathbb{T})}$ as (X, Y), and study the CLT of the following matrix:

$$\mathcal{Q} := \sqrt{n}\mathscr{V}^\top \left[G(X,Y) - \Pi\right]\mathscr{V}, \quad \mathscr{V} := \mathscr{V}_0 \begin{pmatrix} 0 & I_{2r} \\ I_{2r} & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{V}_1 & 0 & 0 & 0 \\ 0 & \mathbf{V}_2 & 0 & 0 \\ 0 & 0 & \mathbf{V}_3 & 0 \\ 0 & 0 & 0 & \mathbf{V}_4 \end{pmatrix},$$

under the conditions (A.1) and (A.2). Since $|\mathbb{T}| \lesssim n^{2\tau_0}$, we have $(n - |\mathbb{T}|)/n = 1 + \mathrm{O}(n^{-1+2\tau_0})$, where $\mathrm{O}(n^{-1+2\tau_0})$ is a negligible error. Hence without loss of generality, we still assume that the dimensions of X and Y are $p \times n$ and $q \times n$ in order to simplify the notations.

For $1 \leqslant a \leqslant 4r$, we denote the a-th column vector of $\mathscr{V}$ by $\mathbf{v}_a$. With the Cramér-Wold device, it suffices to prove that

$$Q_\Lambda := \sqrt{n}\sum_{1\leqslant a\leqslant b\leqslant 4r} \lambda_{ab} Q_{ab} = \sqrt{n}\sum_{a\leqslant b} \lambda_{ab}(G-\Pi)_{\mathbf{v}_a\mathbf{v}_b}$$

is asymptotically Gaussian for any fixed vector of parameters denoted by $\Lambda := (\lambda_{ab})_{a\leqslant b}$. Note that by (3.24), we have the rough bound $|Q_\Lambda| \prec 1$. For our purpose, it suffices to show that the moments of Q_Λ match those of a centered Gaussian random variable asymptotically. This follows immediately from the following claims: (i) the mean of Q_Λ satisfies

$$\mathbb{E}Q_\Lambda = \mathrm{o}(1), \tag{A.4}$$

and (ii) for any fixed integer $k \geqslant 2$, we have that

$$\mathbb{E}Q_\Lambda^k = (k-1)s_\Lambda^2 \mathbb{E}Q_\Lambda^{k-2} + \mathrm{o}(1) \tag{A.5}$$

for some deterministic parameter s_Λ^2. Moreover, the covariances of $\mathcal{Q}$ can be determined from s_Λ^2 as a function of Λ. Once again, to avoid singular behaviours of the resolvents on exceptional low-probability events, we will use the regularized resolvent $\widehat{G}(z)$ in Definition 5.1 with $z = \theta_l + \mathrm{i}n^{-4}$, and prove the CLT for $\widehat{\mathcal{Q}}(z)$ with $G(\theta_l)$ replaced by $\widehat{G}(z)$. The argument in the proof of Claim 5.2 then allows us to show that $\mathcal{Q}(\theta_l)$ satisfies the same asymptotic distribution. However, for simplicity of notations, we still use the notations $G(z)$ and $Q_\Lambda(z)$ in the following proof, while keeping in mind that the bound (5.12) holds for all resolvents in the proof, and hence Lemma 3.2 (iii) can be applied without worry.

Our main tool for the proof of (A.4) and (A.5) is the conventional Gaussian integration by parts. Using the identity $\widehat{H}G = I$ and equation (3.19), we get that

$$G - \Pi = \Pi\left(\Pi^{-1} - \widehat{H}\right)G = \Pi \begin{bmatrix} -(m_{3c} + zn^{-10})I_p & 0 & -X & 0 \\ 0 & -(m_{4c} + zn^{-10})I_q & 0 & -Y \\ -X^\top & 0 & -m_{1c}I_n & 0 \\ 0 & -Y^\top & 0 & -m_{2c}I_n \end{bmatrix} G. \tag{A.6}$$

We first prove (A.4). With (A.6), we can write that

$$\begin{aligned} \mathbb{E}Q_\Lambda := \sqrt{n}\sum_{a\leqslant b}\lambda_{ab}\mathbb{E}Q_{ab} &= \sqrt{n}\sum_{a\leqslant b}\lambda_{ab}\mathbb{E}\left[\begin{pmatrix} -m_{3c}I_p & 0 & 0 & 0 \\ 0 & -m_{4c}I_q & 0 & 0 \\ 0 & 0 & -m_{1c}I_n & 0 \\ 0 & 0 & 0 & -m_{2c}I_n \end{pmatrix} G\right]_{\mathbf{w}_a\mathbf{v}_b} \\ &\quad - \sqrt{n}\sum_{a\leqslant b}\lambda_{ab}\mathbb{E}\left[\begin{pmatrix} 0 & \begin{pmatrix} X & 0 \\ 0 & Y\end{pmatrix} \\ \begin{pmatrix} X^\top & 0 \\ 0 & Y^\top\end{pmatrix} & 0 \end{pmatrix} G\right]_{\mathbf{w}_a\mathbf{v}_b} + \mathrm{O}(n^{-9}), \end{aligned} \tag{A.7}$$

where we have abbreviated $\mathbf{w}_a := \Pi\,\mathbf{v}_a$. For the sum in line (A.7), we expand it as

$$\begin{aligned} &\mathbb{E}\left[\begin{pmatrix} 0 & \begin{pmatrix} X & 0 \\ 0 & Y\end{pmatrix} \\ \begin{pmatrix} X^\top & 0 \\ 0 & Y^\top\end{pmatrix} & 0 \end{pmatrix} G\right]_{\mathbf{w}_a\mathbf{v}_b} \\ &= -\sqrt{n}\mathbb{E}\sum_{i\in\mathcal{I}_1,\mu\in\mathcal{I}_3} X_{i\mu}\left[\mathbf{w}_a(i)G_{\mu\mathbf{v}_b} + \mathbf{w}_a(\mu)G_{i\mathbf{v}_b}\right] - \sqrt{n}\mathbb{E}\sum_{j\in\mathcal{I}_2,\nu\in\mathcal{I}_4} Y_{j\nu}\left[\mathbf{w}_a(j)G_{\nu\mathbf{v}_b} + \mathbf{w}_a(\nu)G_{j\mathbf{v}_b}\right] \\ &= n^{-1/2}\mathbb{E}\sum_{i\in\mathcal{I}_1,\mu\in\mathcal{I}_3}\mathbf{w}_a(i)\left[G_{\mu\mu}G_{i\mathbf{v}_b} + G_{\mu i}G_{\mu\mathbf{v}_b}\right] + n^{-1/2}\mathbb{E}\sum_{i\in\mathcal{I}_1,\mu\in\mathcal{I}_3}\mathbf{w}_a(\mu)\left[G_{ii}G_{\mu\mathbf{v}_b} + G_{i\mu}G_{i\mathbf{v}_b}\right] \\ &\quad + n^{-1/2}\mathbb{E}\sum_{j\in\mathcal{I}_2,\nu\in\mathcal{I}_4}\mathbf{w}_a(j)\left[G_{\nu\nu}G_{j\mathbf{v}_b} + G_{\nu j}G_{\nu\mathbf{v}_b}\right] + n^{-1/2}\mathbb{E}\sum_{j\in\mathcal{I}_2,\nu\in\mathcal{I}_4}\mathbf{w}_a(\nu)\left[G_{jj}G_{\nu\mathbf{v}_b} + G_{j\nu}G_{j\mathbf{v}_b}\right], \end{aligned} \tag{A.8}$$

where in the second step we used Gaussian integration by parts with respect to $X_{i\mu}$ and $Y_{j\nu}$,

$$\mathbb{E}X_{i\mu}f(X_{i\mu}) = n^{-1}\mathbb{E}f'(X_{i\mu}), \quad \mathbb{E}Y_{j\nu}f(Y_{j\nu}) = n^{-1}\mathbb{E}f'(Y_{j\nu}),$$

and the identities

$$\frac{\partial G_{\mathbf{u}\,\mathbf{v}}}{\partial X_{i\mu}} = -G_{\mathbf{u}\,i}G_{\mu\,\mathbf{v}} - G_{\mathbf{u}\,\mu}G_{i\,\mathbf{v}}, \quad \frac{\partial G_{\mathbf{u}\,\mathbf{v}}}{\partial Y_{j\nu}} = -G_{\mathbf{u}\,j}G_{\nu\,\mathbf{v}} - G_{\mathbf{u}\,\nu}G_{j\,\mathbf{v}}, \tag{A.9}$$

for any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$. With the notations in (3.9), we can rewrite (A.8) as

$$\begin{aligned}(\text{A.8}) = \sqrt{n}\mathbb{E}&\left[\begin{pmatrix} m_3 I_p & 0 & 0 & 0 \\ 0 & m_4 I_q & 0 & 0 \\ 0 & 0 & m_1 I_n & 0 \\ 0 & 0 & 0 & m_2 I_n \end{pmatrix} G\right]_{\mathbf{w}_a\,\mathbf{v}_b} \\ &+ n^{-1/2}\mathbb{E}\left[\langle \mathbf{w}_a, J_1 G J_3 G\,\mathbf{v}_b\rangle + \langle \mathbf{w}_a, J_3 G J_1 G\,\mathbf{v}_b\rangle\right] + n^{-1/2}\mathbb{E}\left[\langle \mathbf{w}_a, J_2 G J_4 G\,\mathbf{v}_b\rangle + \langle \mathbf{w}_a, J_4 G J_2 G\,\mathbf{v}_b\rangle\right],\end{aligned} \tag{A.10}$$

where we define the matrices J_α as the block identity matrices,

$$(J_\alpha)_{\mathfrak{a}\mathfrak{b}} = \delta_{\mathfrak{a}\mathfrak{b}}\mathbf{1}_{\mathfrak{a}\in\mathcal{I}_\alpha}, \quad \alpha = 1,2,3,4. \tag{A.11}$$

We claim that

$$\max_{\alpha=1}^{4} |m_\alpha(z) - m_{\alpha c}(z)| \prec n^{-2/3}, \tag{A.12}$$

whose proof will be postponed until we complete the proof of Lemma 4.5. Moreover, $GJ_\alpha G$, $\alpha = 1,2,3,4$, satisfy the anisotropic local laws in Theorem A.1 below, which implies that for any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$,

$$|\langle \mathbf{u}, GJ_\alpha G\,\mathbf{v}\rangle| = \mathrm{O}_\prec(1), \quad \alpha = 1,2,3,4. \tag{A.13}$$

Now plugging (A.10) into (A.7) and using (A.12) and (A.13), we obtain that

$$\mathbb{E}Q_\Lambda = \mathrm{O}_\prec(n^{-1/6}), \tag{A.14}$$

which concludes (A.4).

It remains to prove (A.5). With (A.6), we expand $\mathbb{E}Q_\Lambda^k$ as

$$\mathbb{E}Q_\Lambda^k = \mathbb{E}\sqrt{n}\sum_{a\leqslant b}\lambda_{ab}\mathbb{E}\left[\Pi\begin{pmatrix} -m_{3c}I_p & 0 & -X & 0 \\ 0 & -m_{4c}I_q & 0 & -Y \\ -X^\top & 0 & -m_{1c}I_n & 0 \\ 0 & -Y^\top & 0 & -m_{2c}I_n \end{pmatrix} G\right]_{\mathbf{v}_a\,\mathbf{v}_b} Q_\Lambda^{k-1} + \mathrm{O}(n^{-9})$$

$$= \sqrt{n}\sum_{a\leqslant b}\lambda_{ab}\mathbb{E}\left[\begin{pmatrix} -m_{3c}I_p & 0 & 0 & 0 \\ 0 & -m_{4c}I_q & 0 & 0 \\ 0 & 0 & -m_{1c}I_n & 0 \\ 0 & 0 & 0 & -m_{2c}I_n \end{pmatrix} G\right]_{\mathbf{w}_a\,\mathbf{v}_b} Q_\Lambda^{k-1} + \mathrm{O}(n^{-9}) \tag{A.15}$$

$$- \sqrt{n}\mathbb{E}\sum_{a\leqslant b}\lambda_{ab}\sum_{i\in\mathcal{I}_1,\mu\in\mathcal{I}_3}\mathbf{w}_a(i)X_{i\mu}G_{\mu\,\mathbf{v}_b}Q_\Lambda^{k-1} - \sqrt{n}\mathbb{E}\sum_{a\leqslant b}\lambda_{ab}\sum_{j\in\mathcal{I}_2,\nu\in\mathcal{I}_4}\mathbf{w}_a(j)Y_{j\nu}G_{\nu\,\mathbf{v}_b}Q_\Lambda^{k-1} \tag{A.16}$$

$$- \sqrt{n}\mathbb{E}\sum_{a\leqslant b}\lambda_{ab}\sum_{i\in\mathcal{I}_1,\mu\in\mathcal{I}_3}\mathbf{w}_a(\mu)X_{i\mu}G_{i\,\mathbf{v}_b}Q_\Lambda^{k-1} - \sqrt{n}\mathbb{E}\sum_{a\leqslant b}\lambda_{ab}\sum_{j\in\mathcal{I}_2,\nu\in\mathcal{I}_4}\mathbf{w}_a(\nu)Y_{j\nu}G_{j\,\mathbf{v}_b}Q_\Lambda^{k-1}. \tag{A.17}$$

Then we apply Gaussian integration by parts to the terms in (A.16) and (A.17). First, as we have seen in the $k=1$ case, the terms containing $\partial_{X_{i\mu}}G_{\mu\,\mathbf{v}_b}$, $\partial_{X_{i\mu}}G_{i\,\mathbf{v}_b}$, $\partial_{Y_{j\nu}}G_{\nu\,\mathbf{v}_b}$ and $\partial_{Y_{j\nu}}G_{j\,\mathbf{v}_b}$ will cancel the first term in (A.15), leaving an error of order $\mathrm{O}_\prec(n^{-1/6})$ as in (A.14). Thus we get that

$$\mathbb{E}Q_\Lambda^k = -n^{-1/2}\sum_{a\leqslant b}\lambda_{ab}\mathbb{E}\sum_{i\in\mathcal{I}_1,\mu\in\mathcal{I}_3,}\mathbf{w}_a(i)G_{\mu\,\mathbf{v}_b}\frac{\partial Q_\Lambda^{k-1}}{\partial X_{i\mu}} - n^{-1/2}\sum_{a\leqslant b}\lambda_{ab}\mathbb{E}\sum_{j\in\mathcal{I}_2,\nu\in\mathcal{I}_4}\mathbf{w}_a(j)G_{\nu\,\mathbf{v}_b}\frac{\partial Q_\Lambda^{k-1}}{\partial Y_{j\nu}}$$

$$- n^{-1/2}\sum_{a\leqslant b}\lambda_{ab}\mathbb{E}\sum_{i\in\mathcal{I}_1,\mu\in\mathcal{I}_3}\mathbf{w}_a(\mu)G_{i\,\mathbf{v}_b}\frac{\partial Q_\Lambda^{k-1}}{\partial X_{i\mu}} - n^{-1/2}\sum_{a\leqslant b}\lambda_{ab}\mathbb{E}\sum_{j\in\mathcal{I}_2,\nu\in\mathcal{I}_4}\mathbf{w}_a(\nu)G_{j\,\mathbf{v}_b}\frac{\partial Q_\Lambda^{k-1}}{\partial Y_{j\nu}} + \mathrm{O}_\prec(n^{-1/6})$$

$$= -(k-1)\sum_{a\leqslant b,a'\leqslant b'}\lambda_{ab}\lambda_{a'b'}\mathbb{E}\sum_{i\in\mathcal{I}_1,\mu\in\mathcal{I}_3}\mathbf{w}_a(i)G_{\mu\,\mathbf{v}_b}\frac{\partial G_{\mathbf{v}_{a'}\,\mathbf{v}_{b'}}}{\partial X_{i\mu}}Q_\Lambda^{k-2} \tag{A.18}$$

$$-(k-1)\sum_{a\leqslant b,a'\leqslant b'}\lambda_{ab}\lambda_{a'b'}\mathbb{E}\sum_{j\in\mathcal{I}_2,\nu\in\mathcal{I}_4}\mathbf{w}_a(j)G_{\nu\,\mathbf{v}_b}\frac{\partial G_{\mathbf{v}_{a'}\,\mathbf{v}_{b'}}}{\partial Y_{j\nu}}Q_\Lambda^{k-2} \tag{A.19}$$

$$-(k-1)\sum_{a\leqslant b,a'\leqslant b'}\lambda_{ab}\lambda_{a'b'}\mathbb{E}\sum_{i\in\mathcal{I}_1,\mu\in\mathcal{I}_3}\mathbf{w}_a(\mu)G_{i\,\mathbf{v}_b}\frac{\partial G_{\mathbf{v}_{a'}\,\mathbf{v}_{b'}}}{\partial X_{i\mu}}Q_\Lambda^{k-2} \tag{A.20}$$

$$-(k-1)\sum_{a\leqslant b,a'\leqslant b'}\lambda_{ab}\lambda_{a'b'}\mathbb{E}\sum_{j\in\mathcal{I}_2,\nu\in\mathcal{I}_4}\mathbf{w}_a(\nu)G_{j\,\mathbf{v}_b}\frac{\partial G_{\mathbf{v}_{a'}\,\mathbf{v}_{b'}}}{\partial Y_{j\nu}}Q_\Lambda^{k-2} + \mathrm{O}_\prec(n^{-1/6}). \tag{A.21}$$

To calculate the terms (A.18)–(A.21), we need to use the anisotropic local laws of $GJ_\alpha G$, $\alpha = 1,2,3,4$. We first define the deterministic matrix limits of $GJ_\alpha G$:

$$\Gamma^{(\alpha)}(z) := \begin{bmatrix} \begin{pmatrix} \gamma_1^{(\alpha)}(z)I_p & 0 \\ 0 & \gamma_2^{(\alpha)}(z)I_q \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} \gamma_3^{(\alpha)}(z)I_n & h_\alpha(z)I_n \\ h_\alpha(z)I_n & \gamma_4^{(\alpha)}(z)I_n \end{pmatrix} \end{bmatrix}, \quad \alpha = 1,2,3,4, \tag{A.22}$$

where the γ functions are defined by

$$\begin{aligned}
&\gamma_1^{(1)} := \frac{(1-c_1)^{-1}f_c^2}{m_{3c}^2(f_c^2-t_c^2)}, \quad \gamma_2^{(1)} := \frac{c_2^{-1}t_c^2}{h^2(f_c^2-t_c^2)}, \quad \gamma_3^{(1)} := \frac{(1-c_1)^{-1}f_c^2}{f_c^2-t_c^2} - 1, \quad \gamma_4^{(1)} := \frac{c_2^{-1}m_{4c}^2t_c^2}{h^2(f_c^2-t_c^2)},\\
&\gamma_1^{(2)} := \frac{c_1^{-1}t_c^2}{h^2(f_c^2-t_c^2)}, \quad \gamma_2^{(2)} := \frac{(1-c_2)^{-1}f_c^2}{m_{4c}^2(f_c^2-t_c^2)}, \quad \gamma_3^{(2)} := \frac{c_1^{-1}m_{3c}^2t_c^2}{h^2(f_c^2-t_c^2)}, \quad \gamma_4^{(2)} := \frac{(1-c_2)^{-1}f_c^2}{f_c^2-t_c^2} - 1,\\
&\gamma_1^{(3)} := c_1^{-1}\gamma_3^{(1)}, \quad \gamma_2^{(3)} := c_2^{-1}\gamma_3^{(2)}, \quad \gamma_3^{(3)} := c_1^{-1}m_{3c}^2\gamma_3^{(1)}, \quad \gamma_4^{(3)} := \frac{c_1^{-1}c_2^{-1}h^2t_c^2f_c^2}{f_c^2-t_c^2},\\
&\gamma_1^{(4)} := c_1^{-1}\gamma_4^{(1)}, \quad \gamma_2^{(4)} := c_2^{-1}\gamma_4^{(2)}, \quad \gamma_3^{(4)} := \gamma_4^{(3)}, \quad \gamma_4^{(4)} := c_2^{-1}m_{4c}^2\gamma_4^{(2)}.
\end{aligned} \tag{A.23}$$

On the other hand, the functions h_α are defined by

$$h_\alpha(z) := z^{1/2}h^2(z)\left\{c_1\gamma_1^{(\alpha)}(z)\left[1+(1-z)m_{2c}(z)\right] + c_2\gamma_2^{(\alpha)}(z)\left[1+(1-z)m_{1c}(z)\right]\right\}.$$

Here we recall that t_c is defined in (1.3), $m_{\alpha c}$, $\alpha = 1,2,3,4$, are defined in (3.10)–(3.13), h is defined in (3.17), and f_c is defined in (3.35).

Theorem A.1. *Suppose Assumption 2.1 holds. For any deterministic unit vectors $\mathbf{u},\mathbf{v}\in\mathbb{C}^{\mathcal{I}}$, we have that*

$$\langle \mathbf{u}, G(\theta_l)J_\alpha G(\theta_l)\,\mathbf{v}\rangle - \langle \mathbf{u}, \Gamma^{(\alpha)}(\theta_l)\,\mathbf{v}\rangle \prec n^{-1/2}. \tag{A.24}$$

The proof of Theorem A.1 will be given in Appendix C. Again by the argument in the proof of Claim 5.2, (A.24) also holds for $z = \theta_l + \mathrm{i}n^{-4}$. Now we use this estimate to calculate (A.18)–(A.21) term by term. First for (A.18), using (A.9) we get that

$$\begin{aligned}
-\mathbb{E}\sum_{i\in\mathcal{I}_1,\mu\in\mathcal{I}_3}\mathbf{w}_a(i)G_{\mu\,\mathbf{v}_b}\frac{\partial G_{\mathbf{v}_{a'}\,\mathbf{v}_{b'}}}{\partial X_{i\mu}}Q_\Lambda^{k-2} &= \mathbb{E}(GJ_3G)_{\mathbf{v}_{b'}\,\mathbf{v}_b}\langle \mathbf{v}_a, \Pi J_1 G\,\mathbf{v}_{a'}\rangle Q_\Lambda^{k-2}\\
&+ \mathbb{E}(GJ_3G)_{\mathbf{v}_{a'}\,\mathbf{v}_b}\langle \mathbf{v}_a, \Pi J_1 G\,\mathbf{v}_{b'}\rangle Q_\Lambda^{k-2}.
\end{aligned} \tag{A.25}$$

Now using the local law (3.24), (A.1) and the first equation in (3.14), we can calculate that

$$\langle \mathbf{v}_a, \Pi J_1 G\, \mathbf{v}_{a'}\rangle = c_1 \left(c_1^{-1} m_{1c}\right)^2 \delta_{aa'} \mathbf{1}_{1\leqslant a\leqslant r} + \mathrm{O}_\prec(n^{-1/2}) = c_1 m_{3c}^{-2} \delta_{aa'} \mathbf{1}_{1\leqslant a\leqslant r} + \mathrm{O}_\prec(n^{-1/2}). \tag{A.26}$$

Moreover, using (A.1), (A.2) and the local law for GJ_3G in Theorem A.1, we get that

$$(GJ_3G)_{\mathbf{v}_{b'}\,\mathbf{v}_b} = c_{\alpha(b)} \gamma^{(3)}_{\alpha(b)} \delta_{bb'} + \mathrm{O}_\prec(n^{-1/2+2\tau_0}), \tag{A.27}$$

where we used the notation

$$\alpha(b) := k \quad \text{if} \;\; (k-1)r+1 \leqslant b \leqslant kr, \quad k = 1,2,3,4,$$

and let $c_k \equiv 1$ for $k = 3, 4$. Plugging (A.26) and (A.27) into (A.25), we get that

$$(\mathrm{A.18}) = (k-1) \sum_{1\leqslant a\leqslant r, a\leqslant b} c_1 c_{\alpha(b)} \frac{\lambda_{ab}^2}{m_{3c}^2} \gamma^{(3)}_{\alpha(b)} (1+\delta_{ab}) \mathbb{E} Q_\Lambda^{k-2} + \mathrm{O}_\prec(n^{-1/2+2\tau_0}). \tag{A.28}$$

Similarly, we can get that

$$(\mathrm{A.19}) = (k-1) \sum_{r+1\leqslant a\leqslant 2r, a\leqslant b} c_2 c_{\alpha(b)} \frac{\lambda_{ab}^2}{m_{4c}^2} \gamma^{(4)}_{\alpha(b)} (1+\delta_{ab}) \mathbb{E} Q_\Lambda^{k-2} + \mathrm{O}_\prec(n^{-1/2+2\tau_0}). \tag{A.29}$$

For (A.20), we have that

$$\begin{aligned} -\mathbb{E} \sum_{i\in\mathcal{I}_1, \mu\in\mathcal{I}_3} \mathbf{w}_a(\mu) G_{i\,\mathbf{v}_b} \frac{\partial G_{\mathbf{v}_{a'}\,\mathbf{v}_{b'}}}{\partial X_{i\mu}} Q_\Lambda^{k-2} = \mathbb{E} &\sum_{i\in\mathcal{I}_1, \mu\in\mathcal{I}_3} (GJ_1G)_{\mathbf{v}_{b'}\,\mathbf{v}_b} \langle \mathbf{v}_a, \Pi J_3 G\, \mathbf{v}_{a'}\rangle Q_\Lambda^{k-2} \\ + \mathbb{E} &\sum_{i\in\mathcal{I}_1, \mu\in\mathcal{I}_3} (GJ_1G)_{\mathbf{v}_{a'}\,\mathbf{v}_b} \langle \mathbf{v}_a, \Pi J_3 G\, \mathbf{v}_{b'}\rangle Q_\Lambda^{k-2}. \end{aligned} \tag{A.30}$$

Using (3.24) and (A.2), we get that

$$\langle \mathbf{v}_a, \Pi J_3 G\, \mathbf{v}_{a'}\rangle = m_{3c}^2 \delta_{aa'} \mathbf{1}_{2r+1\leqslant a\leqslant 3r} + h^2 \delta_{aa'} \mathbf{1}_{3r+1\leqslant a\leqslant 4r} + \mathrm{O}_\prec(n^{-1/2+2\tau_0}). \tag{A.31}$$

Using the local law for GJ_1G in Theorem A.1 and (A.2), we get that

$$(GJ_1G)_{\mathbf{v}_{b'}\,\mathbf{v}_b} = \gamma^{(1)}_{\alpha(b)} \delta_{bb'} + \mathrm{O}_\prec(n^{-1/2+2\tau_0}), \quad \text{for} \quad \alpha(b) = 3, 4. \tag{A.32}$$

Plugging (A.31) and (A.32) into (A.30) gives that

$$\begin{aligned} (\mathrm{A.20}) = (k-1) &\sum_{2r+1\leqslant a\leqslant 3r, a\leqslant b} \lambda_{ab}^2 m_{3c}^2 \gamma^{(1)}_{\alpha(b)} (1+\delta_{ab}) \mathbb{E} Q_\Lambda^{k-2} \\ + (k-1) &\sum_{3r+1\leqslant a\leqslant 4r, a\leqslant b} \lambda_{ab}^2 h^2 \gamma^{(1)}_{\alpha(b)} (1+\delta_{ab}) \mathbb{E} Q_\Lambda^{k-2} + \mathrm{O}_\prec(n^{-1/2+2\tau_0}). \end{aligned} \tag{A.33}$$

Similarly, we can get that

$$\begin{aligned} (\mathrm{A.21}) = (k-1) &\sum_{2r+1\leqslant a\leqslant 3r, a\leqslant b} \lambda_{ab}^2 h^2 \gamma^{(2)}_{\alpha(b)} (1+\delta_{ab}) \mathbb{E} Q_\Lambda^{k-2} \\ + (k-1) &\sum_{3r+1\leqslant a\leqslant 4r, a\leqslant b} \lambda_{ab}^2 m_{4c}^2 \gamma^{(2)}_{\alpha(b)} (1+\delta_{ab}) \mathbb{E} Q_\Lambda^{k-2} + \mathrm{O}_\prec(n^{-1/2+\tau_0}). \end{aligned} \tag{A.34}$$

Combining (A.28), (A.29), (A.33) and (A.34), we obtain that

$$\mathbb{E}Q_\Lambda^k = (k-1)s_\Lambda^2 \mathbb{E}Q_\Lambda^{k-2} + \mathrm{O}_\prec(n^{-1/6}),$$

where s_Λ^2 is a function of Λ defined by

$$\begin{aligned} s_\Lambda^2 := & \sum_{1\leqslant a\leqslant r, a\leqslant b} c_1 c_{\alpha(b)} \frac{\lambda_{ab}^2}{m_{3c}^2}\gamma_{\alpha(b)}^{(3)}(1+\delta_{ab}) + \sum_{r+1\leqslant a\leqslant 2r, a\leqslant b} c_2 c_{\alpha(b)} \frac{\lambda_{ab}^2}{m_{4c}^2}\gamma_{\alpha(b)}^{(4)}(1+\delta_{ab}) \\ & + \sum_{2r+1\leqslant a\leqslant 3r, a\leqslant b} \lambda_{ab}^2 \left(m_{3c}^2 \gamma_{\alpha(b)}^{(1)} + h^2 \gamma_{\alpha(b)}^{(2)} \right)(1+\delta_{ab}) \\ & + \sum_{3r+1\leqslant a\leqslant 4r, a\leqslant b} \lambda_{ab}^2 \left(h^2 \gamma_{\alpha(b)}^{(1)} + m_{4c}^2 \gamma_{\alpha(b)}^{(2)} \right)(1+\delta_{ab}). \end{aligned}$$

This concludes (A.5). Combining (A.4) and (A.5), we have shown that $\mathcal{Q}_\Lambda$ is asymptotically Gaussian with zero mean, which indicates that $\mathcal{Q}$ converges weakly to a centered Gaussian matrix by the Cramér-Wold device. Using the definitions of $\gamma_\beta^{(\alpha)}$, $\alpha, \beta = 1,2,3,4$, in (A.23), we obtain from s_Λ^2 that

$$\sqrt{n}\mathcal{Q} \to \begin{pmatrix} b_{11}g_{11} & b_{12}g_{12} & b_{13}g_{13} & b_{14}g_{14} \\ b_{21}g_{21} & b_{22}g_{22} & b_{23}g_{23} & b_{24}g_{24} \\ b_{31}g_{31} & b_{32}g_{32} & b_{33}g_{33} & b_{34}g_{34} \\ b_{41}g_{41} & b_{42}g_{42} & b_{43}g_{43} & b_{44}g_{44} \end{pmatrix}. \tag{A.35}$$

Here $g_{\alpha\beta}$ are Gaussian matrices as defined in Lemma 4.5, and through direct calculations, we can check that $b_{\alpha\beta}$ are given by

$$\begin{aligned} & b_{11} = a_{33}, \quad b_{12} = b_{21} = a_{34}, \quad b_{13} = b_{31} = a_{13}, \quad b_{14} = b_{41} = a_{23}, \quad b_{22} = a_{44}, \\ & b_{23} = b_{32} = a_{14}, \quad b_{24} = b_{42} = a_{24}, \quad b_{33} = a_{11}, \quad b_{34} = b_{43} = a_{12}, \quad b_{44} = a_{22}. \end{aligned} \tag{A.36}$$

In the above calculations, we also used that for $z = \theta_l + \mathrm{i}n^{-4}$,

$$f_c(z) = \frac{m_{3c}(z)m_{4c}(z)}{h^2(z)} = t_l + \mathrm{O}(n^{-4}).$$

Finally, combining (A.35) with (A.3), we can obtain the asymptotic distribution in (4.19), upon renaming the matrices $g_{\alpha\beta}$ and the coefficients $b_{\alpha\beta}$. This concludes Lemma 4.5.

Finally, we give the proof of (A.12).

Proof of (A.12). Recall that we have chosen $z = \theta_l + \mathrm{i}\eta$ with $\eta = n^{-4}$. In the following proof, we denote $z_0 := \theta_l + \mathrm{i}\eta_0$ with $\eta_0 = n^{-2/3}$. By the averaged local law (3.27), we have

$$|m_\alpha(z_0) - m_{\alpha c}(z_0)| \prec n^{-2/3}, \quad \alpha = 1,2,3,4, \tag{A.37}$$

where we also used that $\kappa = |\theta_l - \lambda_+| \sim 1$ due to (2.21). Thus to show (A.12), it suffices to prove that

$$|m_{\alpha c}(z) - m_{\alpha c}(z_0)| \prec n^{-2/3}, \tag{A.38}$$

and

$$|m_\alpha(z) - m_\alpha(z_0)| \prec n^{-2/3}. \tag{A.39}$$

The estimate (A.38) follows directly from the definitions in (3.10)–(3.13). It remains to prove (A.39). We only show the proof for $\alpha = 1$, and all the other cases can be proved in exactly the same way. Using (3.9), (6.43), (6.44) and (6.47), we obtain that

$$|m_1(z) - m_1(z_0)| \prec \frac{\eta_0}{n} \sum_{i=1}^{p} \sum_{k=1}^{p} \left| \left\langle \mathbf{e}_i, S_{xx}^{-1/2} \xi_k \right\rangle \right|^2 = \frac{\eta_0}{n} \operatorname{Tr}(S_{xx}^{-1}) \prec \eta_0 = n^{-2/3},$$

where $\mathbf{e}_i$ is the standard unit vector along the i-th direction and we used (6.10) and (6.11) in the last step. ☐

B Proof of Lemma 4.6

In this section, we provide a proof of Lemma 4.6 using cumulant expansions. With a slight abuse of notation, we consider the following $r \times r$ matrix

$$Q := \sqrt{n} U^\top Y V + \sqrt{n} V^\top Y^\top U + \sqrt{n} O^\top (1 - \mathbb{E})(YY^\top) O,$$

where Y is a $\rho \times n$ random matrix with i.i.d. entries satisfying (2.1) and (2.8), U and O are two $\rho \times r$ deterministic matrices satisfying $\|U\| \leqslant 1$ and $\|O\| \leqslant 1$, and V is an $n \times r$ deterministic matrix satisfying $\|V\| \leqslant 1$ and

$$\|V\|_{\max} \leqslant n^{-c} \tag{B.1}$$

for some constant $0 < c < 1/2$. Moreover, we assume that $r = \mathrm{O}(1)$ and $\rho = \mathrm{O}(n^\tau)$ for a small enough constant $\tau > 0$. Then we claim that Q is asymptotically Gaussian with zero mean. Note that the items (i)–(iv) of Lemma 4.6 all follow from this general claim. In particular, if the entries of Y are i.i.d. Gaussian, then the condition (B.1) is not necessary, because we can rotate V as $YV \mapsto (YO_n)(O_n^\top V)$, where the orthogonal matrix O_n is chosen such that (B.1) holds for $O_n^\top V$ and the distribution of Y is unchanged: $YO_n \overset{d}{=} Y$.

It is trivial to see that $\mathbb{E}Q = 0$. To show that Q is asymptotically Gaussian, with the Cramér-Wold device, we need to prove that

$$Q_\Lambda := \sum_{a \leqslant b} \lambda_{ab} Q_{ab}$$

is asymptotically Gaussian for any fixed vector of parameters denoted by $\Lambda := (\lambda_{ab})_{a \leqslant b}$. For this purpose, we use Stein's method [34], i.e. we will show that for any $f \in C_c^\infty(\mathbb{R})$,

$$\mathbb{E} Q_\Lambda f(Q_\Lambda) = s_\Lambda^2 \mathbb{E} f'(Q_\Lambda) + \mathrm{o}(1) \tag{B.2}$$

for some deterministic parameter s_Λ^2. This gives the CLT for $\sqrt{n} \sum_{a \leqslant b} \lambda_{ab} Q_{ab}$, which implies that Q converges weakly to a centered Gaussian matrix, whose covariances can be determined through s_Λ^2.

For simplicity, we denote $X := \sqrt{n} Y$, such that the entries of X are i.i.d. random variables with mean zero and variance one. Moreover, for any fixed $l \in \mathbb{N}$ there is a constant $\mu_l > 0$ such that

$$\mathbb{E}|X_{11}|^l \leqslant \mu_l. \tag{B.3}$$

We will prove (B.2) with the following cumulant expansion formula, whose proof can be found in [30, Proposition 3.1] and [25, Section II].

Lemma B.1. *Let $f \in C^{l+1}(\mathbb{R})$ for some fixed $l \in \mathbb{N}$. Suppose ξ is a centered random variable whose first $l+2$ moments are finite. Let $\kappa_k(\xi)$ be the k-th cumulant of ξ. Then we have that*

$$\mathbb{E}[\xi f(\xi)] = \sum_{k=1}^{l} \frac{\kappa_{k+1}(\xi)}{k!} \mathbb{E} f^{(k)}(\xi) + \mathcal{E}_l, \tag{B.4}$$

where the error term satisfies that for any $\chi > 0$,

$$|\mathcal{E}_l| \leqslant C_l \mathbb{E}\left[|\xi|^{l+2}\right] \sum_{|t|\leqslant \chi} |f^{(l+1)}(t)| + C_l \mathbb{E}\left[|\xi|^{l+2}\mathbf{1}(|\xi| > \chi)\right] \sum_{t\in\mathbb{R}} |f^{(l+1)}(t)|. \tag{B.5}$$

We expand the left-hand side of (B.2) as

$$\begin{aligned}\mathbb{E}Q_\Lambda f(Q_\Lambda) = \mathbb{E}\sum_{a\leqslant b}\lambda_{ab} \sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} X_{i\mu}(U_{ia}V_{\mu b} + U_{ib}V_{\mu a}) f(Q_\Lambda) \\ + \mathbb{E}\sum_{a\leqslant b}\lambda_{ab} \sum_{1\leqslant i,j\leqslant \rho} \frac{1}{\sqrt{n}} \sum_{1\leqslant \mu\leqslant n} (X_{i\mu}X_{j\mu} - \delta_{ij}) O_{ia}O_{jb} f(Q_\Lambda).\end{aligned} \tag{B.6}$$

We first study the first term on the right-hand side of (B.6). For any fixed $a \leqslant b$, we apply the expansion (B.4) with $\xi = X_{i\mu}$ and $l = 2$ to get that

$$\begin{aligned}\sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} U_{ia}V_{\mu b}\mathbb{E}_{X_{i\mu}}\left[X_{i\mu} f(Q_\Lambda)\right] = \sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} U_{ia}V_{\mu b}\mathbb{E}_{X_{i\mu}}\left[\frac{\partial Q_\Lambda}{\partial X_{i\mu}} f'(Q_\Lambda)\right] \\ + \frac{\kappa_3}{2}\sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} U_{ia}V_{\mu b}\mathbb{E}_{X_{i\mu}}\left[2\sum_{a'\leqslant b'}\lambda_{a'b'}\frac{O_{ia'}O_{ib'}}{\sqrt{n}} f'(Q_\Lambda) + \left(\frac{\partial Q_\Lambda}{\partial X_{i\mu}}\right)^2 f''(Q_\Lambda)\right] + \mathcal{E}_2(X_{i\mu}).\end{aligned} \tag{B.7}$$

Here $\kappa_3 \equiv \kappa_3(X_{i\mu})$ is the third cumulant of $X_{i\mu}$, $\mathcal{E}_2(X_{i\mu})$ satisfies (B.5) with the function $f(Q_\Lambda(X_{i\mu}))$, and

$$\frac{\partial Q_\Lambda}{\partial X_{i\mu}} = \sum_{a'\leqslant b'}\lambda_{a'b'}\left[(U_{ia'}V_{\mu b'} + U_{ib'}V_{\mu a'}) + \sum_{1\leqslant j\leqslant \rho}\frac{1}{\sqrt{n}} X_{j\mu}(O_{ia'}O_{jb'} + O_{ja'}O_{ib'})\right] \prec n^{-c}, \tag{B.8}$$

where we used (B.1) in the second step. The expectation of the first term on the right-hand side of (B.7) is

$$\begin{aligned}&\mathbb{E}\sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} U_{ia}V_{\mu b}\frac{\partial Q_\Lambda}{\partial X_{i\mu}} f'(Q_\Lambda) \\ &= \sum_{a'\leqslant b'}\lambda_{a'b'}\sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} U_{ia}V_{\mu b}(U_{ia'}V_{\mu b'} + U_{ib'}V_{\mu a'})\mathbb{E}f'(Q_\Lambda) + \mathrm{O}_\prec(n^{-1/2+\tau}),\end{aligned} \tag{B.9}$$

where we used Lemma 4.3 to bound that

$$\Big|\sum_{1\leqslant \mu\leqslant n} n^{-1/2}V_{\mu b}X_{j\mu}\Big| \prec n^{-1/2}\Big(\sum_\mu |V_{\mu b}|^2\Big)^{1/2} \leqslant n^{-1/2}. \tag{B.10}$$

Next using (B.8) and $\rho = \mathrm{O}(n^\tau)$, we can bound that

$$\begin{aligned}\mathbb{E}\sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} U_{ia}V_{\mu b}\left(\frac{\partial Q_\Lambda}{\partial X_{i\mu}}\right)^2 f''(Q_\Lambda) \prec n^{-c}\sum_{a'\leqslant b'}\sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} |U_{ia}||V_{\mu b}|(|U_{ia'}||V_{\mu b'}| + |U_{ib'}||V_{\mu a'}|) \\ + n^{-c}\sum_{a'\leqslant b'}\sum_{1\leqslant i,j\leqslant \rho, 1\leqslant \mu\leqslant n} |U_{ia}||V_{\mu b}|\frac{1}{\sqrt{n}} = \mathrm{O}(n^{-c+\tau/2}),\end{aligned} \tag{B.11}$$

where we used Cauchy-Schwarz inequality in the second step. Finally, we bound $\mathcal{E}_2$ by taking $\chi = n^\varepsilon$ for a small constant $\varepsilon > 0$. We need to bound

$$\frac{\partial^3 f(Q_\Lambda)}{\partial X_{i\mu}^3} = 4\sum_{a'\leqslant b'}\lambda_{a'b'}\frac{O_{ia'}O_{ib'}}{\sqrt{n}}\frac{\partial Q_\Lambda}{\partial X_{i\mu}} f''(Q_\Lambda) + \left(\frac{\partial Q_\Lambda}{\partial X_{i\mu}}\right)^3 f'''(Q_\Lambda).$$

Using the compact support condition of f, it is easy to check that

$$\sup_{|X_{i\mu}|\leqslant n^{\varepsilon}} \left|\frac{\partial^3 f(Q_\Lambda)}{\partial X_{i\mu}^3}\right| \lesssim \sum_{a\leqslant b} \frac{1}{\sqrt{n}} \left(\frac{n^{\varepsilon}}{\sqrt{n}} + \frac{\sum_{j\neq i}|X_{j\mu}|}{\sqrt{n}} + |V_{\mu a}| + |V_{\mu b}|\right) + \sum_{a\leqslant b} \left(\frac{n^{\varepsilon}}{\sqrt{n}} + \frac{\sum_{j\neq i}|X_{j\mu}|}{\sqrt{n}} + |V_{\mu a}| + |V_{\mu b}|\right)^3,$$

and

$$\sup_{X_{i\mu}\in\mathbb{R}} \left|\frac{\partial^3 f(Q_\Lambda)}{\partial X_{i\mu}^3}\right| = \mathrm{O}(1).$$

On the other hand, applying Markov's inequality to (B.3), we obtain the bound

$$\mathbb{E}\left[|X_{i\mu}|^4 \mathbf{1}(|X_{i\mu}| > n^{\varepsilon})\right] \leqslant n^{-D} \quad \text{for any constant } D > 0.$$

Combining the above three estimates, we obtain that

$$|\mathcal{E}_2(X_{i\mu})| \lesssim \mathbb{E} \sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} |U_{ia}|\,|V_{\mu b}| \sum_{a'\leqslant b'} \frac{1}{\sqrt{n}} \left(\frac{n^{\varepsilon}}{\sqrt{n}} + \frac{\sum_{j\neq i}|X_{j\mu}|}{\sqrt{n}} + |V_{\mu a'}| + |V_{\mu b'}|\right)$$
$$+ \mathbb{E} \sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} |U_{ia}|\,|V_{\mu b}| \sum_{a'\leqslant b'} \left(\frac{n^{\varepsilon}}{\sqrt{n}} + \frac{\sum_{j\neq i}|X_{j\mu}|}{\sqrt{n}} + |V_{\mu a'}| + |V_{\mu b'}|\right)^3 + n^{-D} \lesssim n^{-2c+\tau/2}, \quad \text{(B.12)}$$

where we used (B.1) in the second step. Now plugging (B.9), (B.11) and (B.12) into (B.7), we obtain that

$$\mathbb{E} \sum_{1\leqslant i\leqslant r, 1\leqslant \mu\leqslant n} U_{ia}V_{\mu b}X_{i\mu}f(Q_\Lambda) = \sum_{a'\leqslant b'} \lambda_{a'b'} \sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} U_{ia}V_{\mu b}(U_{ia'}V_{\mu b'} + U_{ib'}V_{\mu a'})\mathbb{E}f'(Q_\Lambda)$$
$$+\kappa_3 \sum_{a'\leqslant b'} \lambda_{a'b'} \sum_{1\leqslant i\leqslant \rho, 1\leqslant \mu\leqslant n} U_{ia}V_{\mu b}\frac{O_{ia'}O_{ib'}}{\sqrt{n}}\mathbb{E}f'(Q_\Lambda) + \mathrm{O}(n^{-c+\tau/2}). \quad \text{(B.13)}$$

Then we calculate the second term on the right-hand side of (B.6). For any fixed $a \leqslant b$, we need to study

$$\sum_{1\leqslant i,j\leqslant \rho, 1\leqslant \mu\leqslant n} \frac{1}{\sqrt{n}} O_{ia}O_{jb}\mathbb{E}_{X_{i\mu}}\left[(X_{i\mu}X_{j\mu} - \delta_{ij})f(Q_\Lambda)\right].$$

We only consider the hardest case with $i = j$, and the $i \neq j$ case can be handled in a similar way. For any fixed $1 \leqslant i \leqslant \rho$, we apply the expansion (B.4) with $\xi = X_{i\mu}$ and $l = 3$ to get that

$$\sum_{1\leqslant \mu\leqslant n} \frac{1}{\sqrt{n}}\mathbb{E}_{X_{i\mu}}\left[X_{i\mu}X_{i\mu}f(Q_\Lambda) - f(Q_\Lambda)\right] = \frac{1}{\sqrt{n}} \sum_{1\leqslant \mu\leqslant n} \mathbb{E}_{X_{i\mu}} X_{i\mu}\frac{\partial Q_\Lambda}{\partial X_{i\mu}} f'(Q_\Lambda)$$
$$+ \frac{\kappa_3}{2\sqrt{n}} \sum_{1\leqslant \mu\leqslant n} E_{X_{i\mu}} \left[2\frac{\partial Q_\Lambda}{\partial X_{i\mu}} f'(Q_\Lambda) + C_i X_{i\mu} f'(Q_\Lambda) + X_{i\mu}\left(\frac{\partial Q_\Lambda}{\partial X_{i\mu}}\right)^2 f''(Q_\Lambda)\right]$$
$$+ \frac{\kappa_4}{6\sqrt{n}} \sum_{1\leqslant \mu\leqslant n} E_{X_{i\mu}} \left[3C_i f'(Q_\Lambda) + 3\left(\frac{\partial Q_\Lambda}{\partial X_{i\mu}}\right)^2 f''(Q_\Lambda) + 3C_i X_{i\mu}\frac{\partial Q_\Lambda}{\partial X_{i\mu}} f''(Q_\Lambda)\right]$$

$$+ \frac{\kappa_4}{6\sqrt{n}} \sum_{1 \leqslant \mu \leqslant n} E_{X_{i\mu}} \left[X_{i\mu} \left(\frac{\partial Q_\Lambda}{\partial X_{i\mu}} \right)^3 f'''(Q_\Lambda) \right] + \mathcal{E}_3(X_{i\mu}), \tag{B.14}$$

where $\kappa_4 \equiv \kappa_4(X_{i\mu})$ is the fourth cumulant of $X_{i\mu}$, $\mathcal{E}_3(X_{i\mu})$ satisfies (B.5) with the function $X_{i\mu} f(Q_\Lambda(X_{i\mu}))$, and we have abbreviated that

$$C_i := \frac{\partial^2 Q_\Lambda}{\partial X_{i\mu}^2} = 2 \sum_{a' \leqslant b'} \lambda_{a'b'} \frac{O_{ia'} O_{ib'}}{\sqrt{n}} = \mathrm{O}(n^{-1/2}). \tag{B.15}$$

Using (B.8), we can bound that

$$\frac{1}{\sqrt{n}} \sum_{1 \leqslant \mu \leqslant n} \left(\frac{\partial Q_\Lambda}{\partial X_{i\mu}} \right)^2 f''(Q_\Lambda) \prec n^{-c} \sum_{a' \leqslant b'} \lambda_{a'b'} \frac{1}{\sqrt{n}} \sum_{\mu} \left(|V_{\mu a'}| + |V_{\mu b'}| + n^{-1/2+\tau} \right) \lesssim n^{-c+\tau}.$$

Similarly, we can get the bounds

$$\frac{1}{\sqrt{n}} \sum_{1 \leqslant \mu \leqslant n} X_{i\mu} \left(\frac{\partial Q_\Lambda}{\partial X_{i\mu}} \right)^2 f''(Q_\Lambda) \prec n^{-c+\tau}, \quad \frac{1}{\sqrt{n}} \sum_{1 \leqslant \mu \leqslant n} X_{i\mu} \left(\frac{\partial Q_\Lambda}{\partial X_{i\mu}} \right)^3 f'''(Q_\Lambda) \prec n^{-2c+\tau},$$

$$\frac{1}{\sqrt{n}} \sum_{1 \leqslant \mu \leqslant n} C_i X_{i\mu} \frac{\partial Q_\Lambda}{\partial X_{i\mu}} f''(Q_\Lambda) \prec n^{-1/2+\tau}.$$

On the other hand, with Lemma 4.3, we can obtain the estimates

$$\frac{1}{n} \sum_{\mu} X_{i\mu} X_{j\mu} = \delta_{ij} + \mathrm{O}_\prec(n^{-1/2}), \quad \frac{1}{\sqrt{n}} \sum_{1 \leqslant \mu \leqslant n} X_{i\mu} \prec n^{-1/2}.$$

Using these two estimates and (B.10), we get that

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{1 \leqslant \mu \leqslant n} X_{i\mu} \frac{\partial Q_\Lambda}{\partial X_{i\mu}} f'(Q_\Lambda) &= \sum_{a' \leqslant b'} \lambda_{a'b'} \sum_{1 \leqslant j \leqslant \rho} \left(\frac{1}{n} \sum_{\mu} X_{i\mu} X_{j\mu} \right) (O_{ia'} O_{jb'} + O_{ja'} O_{ib'}) f'(Q_\Lambda) + \mathrm{O}_\prec(n^{-1/2}) \\
&= 2 \sum_{a' \leqslant b'} \lambda_{a'b'} O_{ia'} O_{ib'} f'(Q_\Lambda) + \mathrm{O}_\prec(n^{-1/2}); \\
\frac{1}{\sqrt{n}} \sum_{1 \leqslant \mu \leqslant n} \frac{\partial Q_\Lambda}{\partial X_{i\mu}} f'(Q_\Lambda) &= \sum_{a' \leqslant b'} \lambda_{a'b'} \left[\frac{1}{\sqrt{n}} \sum_{\mu} (U_{ia'} V_{\mu b'} + U_{ib'} V_{\mu a'}) + \sum_{1 \leqslant j \leqslant r} \frac{1}{n} \sum_{\mu} X_{j\mu} (O_{ia'} O_{jb'} + O_{ja'} O_{ib'}) \right] \\
&= \sum_{a' \leqslant b'} \lambda_{a'b'} \frac{1}{\sqrt{n}} \sum_{\mu} (U_{ia'} V_{\mu b'} + U_{ib'} V_{\mu a'}) + \mathrm{O}_\prec(n^{-1/2}); \\
\frac{1}{\sqrt{n}} \sum_{1 \leqslant \mu \leqslant n} C_i X_{i\mu} f'(Q_\Lambda) &= \mathrm{O}_\prec(n^{-1/2}).
\end{aligned}$$

Finally, $\mathcal{E}_3(X_{i\mu})$ can be estimated in a similar way as $\mathcal{E}_2(X_{i\mu})$: $\mathbb{E}\mathcal{E}_3(X_{i\mu}) \leqslant n^{-c}$. We omit the details of its proof. Combining the above estimates and using Lemma 3.2 (iii), we obtain that

$$\sum_{1 \leqslant i,j \leqslant \rho, 1 \leqslant \mu \leqslant n} \frac{1}{\sqrt{n}} O_{ia} O_{jb} \mathbb{E} \left[(X_{i\mu} X_{j\mu} - \delta_{ij}) f(Q_\Lambda) \right] = s_i^2 \mathbb{E} f'(Q_\Lambda) + \mathrm{O}_\prec(n^{-c+2\tau})$$

for a deterministic parameter s_i^2. Combining this equation with (B.13), we obtain (B.2), which concludes Lemma 4.6.

C Proof of Theorem A.1

Finally, in this section we give the proof of Theorem A.1. We first record the following simple estimate, which can be verified through direct calculations using (3.10)–(3.13).

Lemma C.1 (Lemma 3.2 of [40])**.** *Fix any constants $c, C > 0$. If (2.9) holds, then for $z \in \mathbb{C}_+ \cap \{z : c \leqslant |z| \leqslant C\}$ and $\alpha = 1, 2, 3, 4$, the following estimates hold:*

$$|m_{\alpha c}(z)| \sim 1, \quad \left|z^{-1} - (m_{1c}(z) + m_{2c}(z)) + (z-1)m_{1c}(z)m_{2c}(z)\right| \sim 1. \tag{C.1}$$

C.1 Resolvents and limiting laws

We begin the proof by introducing some new resolvents. With the $H(\theta_l)$ in (3.6), we define the following form of generalized resolvents

$$\mathcal{R}(\mathbf{w}) := \left[H(X, Y, \theta_l) - \begin{pmatrix} w_1 I_p & 0 & 0 & 0 \\ 0 & w_2 I_q & 0 & 0 \\ 0 & 0 & w_3 I_n & 0 \\ 0 & 0 & 0 & w_4 I_n \end{pmatrix} \right]^{-1}, \tag{C.2}$$

where $\mathbf{w} = (w_1, w_2, w_3, w_4) \in \mathbb{C}_+^4$ is a new vector of spectral parameters. Then we have the simple identity

$$G J_\alpha G = \left. \frac{\partial \mathcal{R}(\mathbf{w})}{\partial w_\alpha} \right|_{\mathbf{w}=0}. \tag{C.3}$$

Hence to obtain the local laws on $G(\theta_l) J_\alpha G(\theta_l)$, it suffices to study the local law $\mathcal{R}(\mathbf{w})$ for the spectral parameters $\mathbf{w}$ around the origin.

In the following proof, we only prove the local law for GJ_1G, while the proofs for $GJ_\alpha G$ with $\alpha = 2, 3, 4$ are similar. For this purpose, it suffices to use spectral parameters $\mathbf{w}$ with $w_2 = w_3 = w_4 = 0$. With slight abuse of notations, we shall prove a local law for the following resolvent

$$\mathcal{R}(z, z') := \left[H(X, Y, \theta_l) - \begin{pmatrix} z I_p & 0 & 0 & 0 \\ 0 & z' I_q & 0 & 0 \\ 0 & 0 & z' I_n & 0 \\ 0 & 0 & 0 & z' I_n \end{pmatrix} \right]^{-1}, \quad z, z' \in \mathbb{C}_+. \tag{C.4}$$

We denote the $\mathcal{I}_\alpha \times \mathcal{I}_\alpha$ block of $\mathcal{R}$ by $\mathcal{R}_\alpha$ for $\alpha = 1, 2, 3, 4$. Then as in (3.9), we introduce the following averaged partial traces

$$\omega_\alpha(z, z') := \frac{1}{n} \operatorname{Tr} \mathcal{R}_\alpha(z, z') = \frac{1}{n} \sum_{\mathfrak{a} \in \mathcal{I}_\alpha} \mathcal{R}_{\mathfrak{a}\mathfrak{a}}(z, z'), \quad \alpha = 1, 2, 3, 4. \tag{C.5}$$

Since H is symmetric and has real eigenvalues, we immediately obtain the following deterministic bound

$$\left\| \mathcal{R}(z, z') \right\| \leqslant \frac{C}{\min(\operatorname{Im} z, \operatorname{Im} z')}. \tag{C.6}$$

Most of the time we will choose $z' = 0$. But to avoid the singular behaviours of $\mathcal{R}$ on exceptional low-probability events, we sometimes will choose, say $z' = \mathrm{i} n^{-4}$, so that $\|\mathcal{R}(z, z')\| = \mathrm{O}(n^4)$ by (C.6) and hence Lemma 3.2 (iii) can be applied.

We now describe the deterministic limit of $\mathcal{R}(z,0)$. We first define the deterministic limit $(\omega_{\alpha c}(z))_{\alpha=1}^4$ of $(\omega_\alpha(z,0))_{\alpha=1}^4$, as the unique solution to the following system of self-consistent equations

$$\begin{aligned} \frac{c_1}{\omega_{1c}} &= -z-\omega_{3c}, \quad \omega_{3c}=(\theta_l-1)\frac{1+(1-\theta_l)\omega_{2c}}{[1+(1-\theta_l)\omega_{1c}][1+(1-\theta_l)\omega_{2c}]-\theta_l^{-1}}, \\ \frac{c_2}{\omega_{2c}} &= -\omega_{4c}, \quad \omega_{4c}=(\theta_l-1)\frac{1+(1-\theta_l)\omega_{1c}}{[1+(1-\theta_l)\omega_{1c}][1+(1-\theta_l)\omega_{2c}]-\theta_l^{-1}}, \end{aligned} \tag{C.7}$$

such that $\operatorname{Im}\omega_{\alpha c}(z)>0$ whenever $z\in\mathbb{C}_+$. Moreover, we define the function

$$g_1(z) := \frac{(\theta_l-1)\theta_l^{-1/2}}{[1+(1-\theta_l)\omega_{1c}(z)][1+(1-\theta_l)\omega_{2c}(z)]-\theta_l^{-1}}. \tag{C.8}$$

Then the matrix limit of $\mathcal{R}(z,0)$ is defined by

$$\Gamma(z) := \begin{bmatrix} \begin{pmatrix} c_1^{-1}\omega_{1c}(z)I_p & 0 \\ 0 & c_2^{-1}\omega_{2c}(z)I_q \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} \omega_{3c}(z)I_n & g_1(z)I_n \\ g_1(z)I_n & \omega_{4c}(z)I_n \end{pmatrix} \end{bmatrix}. \tag{C.9}$$

The following lemma gives the existence and uniqueness of the solution $(\omega_{\alpha c}(z))_{\alpha=1}^4$.

Lemma C.2. *There exist constants $c_0, C_0>0$ depending only on c_1, c_2 and δ_l in* (2.21) *such that the following statements hold. If $|z|\leqslant c_0$, then there exists a unique solution to* (C.7) *under the condition*

$$\max_{\alpha=1}^4 |\omega_{\alpha c}(z)-m_{\alpha c}(\theta_l)| \leqslant c_0. \tag{C.10}$$

Moreover, the solution satisfies

$$\max_{\alpha=1}^4 |\omega_{\alpha c}(z)-m_{\alpha c}(\theta_l)| \leqslant C_0|z|. \tag{C.11}$$

Proof. The proof is a standard application of the contraction principle. We abbreviate $m_{\alpha c}\equiv m_{\alpha c}(\theta_l)$ and $\varepsilon_\alpha(z):=\omega_{\alpha c}(z)-m_{\alpha c}(\theta_l)$ with $|\varepsilon_\alpha|\leqslant\tilde{c}$ for a sufficiently small constant $\tilde{c}>0$. From (C.7), we obtain the following equations for $(\omega_{1c},\omega_{2c})$:

$$\begin{aligned} \frac{c_1}{\omega_{1c}} &= -z+(1-\theta_l)\frac{1+(1-\theta_l)\omega_{2c}}{[1+(1-\theta_l)\omega_{1c}][1+(1-\theta_l)\omega_{2c}]-\theta_l^{-1}}, \\ \frac{c_2}{\omega_{2c}} &= (1-\theta_l)\frac{1+(1-\theta_l)\omega_{1c}}{[1+(1-\theta_l)\omega_{1c}][1+(1-\theta_l)\omega_{2c}]-\theta_l^{-1}}. \end{aligned} \tag{C.12}$$

On the other hand, using (3.14)–(3.16), we can check that $m_{1c}(\theta_l)$ and $m_{2c}(\theta_l)$ satisfy the following equations:

$$\begin{aligned} \frac{c_1}{m_{1c}(\theta_l)} &= (1-\theta_l)\frac{1+(1-\theta_l)m_{2c}(\theta_l)}{[1+(1-\theta_l)m_{1c}(\theta_l)][1+(1-\theta_l)m_{2c}(\theta_l)]-\theta_l^{-1}}, \\ \frac{c_2}{m_{2c}(\theta_l)} &= (1-\theta_l)\frac{1+(1-\theta_l)m_{1c}(\theta_l)}{[1+(1-\theta_l)m_{1c}(\theta_l)][1+(1-\theta_l)m_{2c}(\theta_l)]-\theta_l^{-1}}. \end{aligned} \tag{C.13}$$

Subtract (C.13) from (C.12), we get that

$$\begin{aligned} \varepsilon_1\frac{c_1}{(m_{1c}+\varepsilon_1)m_{1c}} &= z+(1-\theta_l)^2\frac{g(m_{2c}+\varepsilon_2)g(m_{2c})\varepsilon_1+\theta_l^{-1}\varepsilon_2}{[g(m_{1c}+\varepsilon_1)g(m_{2c}+\varepsilon_2)-\theta_l^{-1}][g(m_{1c})g(m_{2c})-\theta_l^{-1}]}, \\ \varepsilon_2\frac{c_2}{(m_{2c}+\varepsilon_2)m_{2c}} &= (1-\theta_l)^2\frac{g(m_{1c}+\varepsilon_1)g(m_{1c})\varepsilon_2+\theta_l^{-1}\varepsilon_1}{[g(m_{1c}+\varepsilon_1)g(m_{2c}+\varepsilon_2)-\theta_l^{-1}][g(m_{1c})g(m_{2c})-\theta_l^{-1}]}, \end{aligned} \tag{C.14}$$

where we have abbreviated $g(x) := 1 + (1-\theta_l)x$. Inspired by the above equations, we define iteratively a sequence of vectors $\boldsymbol{\varepsilon}^{(k)} = (\varepsilon_1^{(k)}, \varepsilon_2^{(k)}) \in \mathbb{C}^2$ such that $\boldsymbol{\varepsilon}^{(0)} = \mathbf{0} \in \mathbb{C}^2$, and

$$\begin{aligned}
&\left\{\frac{c_1}{m_{1c}^2} - \frac{(1-\theta_l)^2 g(m_{2c})^2}{[g(m_{1c})g(m_{2c}) - \theta_l^{-1}]^2}\right\}\varepsilon_1^{(k+1)} - \frac{(1-\theta_l)^2\theta_l^{-1}}{[g(m_{1c})g(m_{2c}) - \theta_l^{-1}]^2}\varepsilon_2^{(k+1)} = z + \frac{c_1(\varepsilon_1^{(k)})^2}{m_{1c}^2(m_{1c}+\varepsilon_1^{(k)})} \\
&\quad + \frac{(1-\theta_l)^2}{g(m_{1c})g(m_{2c}) - \theta_l^{-1}}\left\{\frac{g(m_{2c}+\varepsilon_2^{(k)})g(m_{2c})\varepsilon_1^{(k)} + \theta_l^{-1}\varepsilon_2^{(k)}}{g(m_{1c}+\varepsilon_1^{(k)})g(m_{2c}+\varepsilon_2^{(k)}) - \theta_l^{-1}} - \frac{g(m_{2c})^2\varepsilon_1^{(k)} + \theta_l^{-1}\varepsilon_2^{(k)}}{g(m_{1c})g(m_{2c}) - \theta_l^{-1}}\right\}, \\
&\left\{\frac{c_2}{m_{2c}^2} - \frac{(1-\theta_l)^2 g(m_{1c})^2}{[g(m_{1c})g(m_{2c}) - \theta_l^{-1}]^2}\right\}\varepsilon_2^{(k+1)} - \frac{(1-\theta_l)^2\theta_l^{-1}}{[g(m_{1c})g(m_{2c}) - \theta_l^{-1}]^2}\varepsilon_1^{(k+1)} = \frac{c_2(\varepsilon_2^{(k)})^2}{m_{2c}^2(m_{2c}+\varepsilon_2^{(k)})} \\
&\quad + \frac{(1-\theta_l)^2}{g(m_{1c})g(m_{2c}) - \theta_l^{-1}}\left\{\frac{g(m_{1c}+\varepsilon_1^{(k)})g(m_{1c})\varepsilon_2^{(k)} + \theta_l^{-1}\varepsilon_1^{(k)}}{g(m_{1c}+\varepsilon_1^{(k)})g(m_{2c}+\varepsilon_2^{(k)}) - \theta_l^{-1}} - \frac{g(m_{1c})^2\varepsilon_2^{(k)} + \theta_l^{-1}\varepsilon_1^{(k)}}{g(m_{1c})g(m_{2c}) - \theta_l^{-1}}\right\}.
\end{aligned} \tag{C.15}$$

In other words, the two equations in (C.15) define a mapping $\mathbf{f} : \ell^\infty(\mathbb{Z}_2) \to \ell^\infty(\mathbb{Z}_2)$, so that

$$\boldsymbol{\varepsilon}^{(k+1)} = \mathbf{f}(\boldsymbol{\varepsilon}^{(k)}), \quad \mathbf{f}(\mathbf{x}) := S^{-1}\begin{pmatrix} z \\ 0 \end{pmatrix} + S^{-1}\mathbf{e}(\mathbf{x}), \tag{C.16}$$

where

$$S := \begin{bmatrix} \frac{c_1}{m_{1c}^2} - \frac{\theta_l^2(1-\theta_l)^2}{(1-t_l)^2}g(m_{2c})^2 & -\frac{(1-\theta_l)^2\theta_l}{(1-t_l)^2} \\ -\frac{(1-\theta_l)^2\theta_l}{(1-t_l)^2} & \frac{c_2}{m_{2c}^2} - \frac{\theta_l^2(1-\theta_l)^2}{(1-t_l)^2}g(m_{1c})^2 \end{bmatrix},$$

and

$$\mathbf{e}(\mathbf{x}) := \begin{bmatrix} \frac{c_1 x_1^2}{m_{1c}^2(m_{1c}+x_1)} - \frac{\theta_l(1-\theta_l)^2}{1-t_l}\left\{\frac{g(m_{2c}+x_2)g(m_{2c})x_1 + \theta_l^{-1}x_2}{g(m_{1c}+x_1)g(m_{2c}+x_2) - \theta_l^{-1}} - \frac{g(m_{2c})^2 x_1 + \theta_l^{-1}x_2}{g(m_{1c})g(m_{2c}) - \theta_l^{-1}}\right\} \\ \frac{c_2 x_2^2}{m_{2c}^2(m_{2c}+x_2)} - \frac{\theta_l(1-\theta_l)^2}{1-t_l}\left\{\frac{g(m_{1c}+x_1)g(m_{1c})x_2 + \theta_l^{-1}x_1}{g(m_{1c}+x_1)g(m_{2c}+x_2) - \theta_l^{-1}} - \frac{g(m_{1c})^2 x_2 + \theta_l^{-1}x_1}{g(m_{1c})g(m_{2c}) - \theta_l^{-1}}\right\} \end{bmatrix}.$$

Here we have used $\theta_l g(m_{1c})g(m_{2c}) = f_c(\theta_l) = t_l$ (which follows from (3.17) and (3.35)) to simplify the expressions a little bit.

With direct calculation, we can check that under (2.21), there exist constants $\widetilde{c}, \widetilde{C} > 0$ depending only on c_1, c_2 and δ_l such that

$$\|S^{-1}\|_{\ell^\infty \to \ell^\infty} \leqslant \widetilde{C}, \quad \text{and} \quad \|\mathbf{e}(\mathbf{x})\|_\infty \leqslant \widetilde{C}\|\mathbf{x}\|_\infty^2 \ \text{ for } \ \|\mathbf{x}\|_\infty \leqslant \widetilde{c}. \tag{C.17}$$

With (C.17), it is easy to check that there exists a sufficiently small constant $\tau > 0$ depending only on $\widetilde{C}$, such that $\mathbf{f}$ is a self-mapping

$$\mathbf{f} : B_r\left(\ell^\infty(\mathbb{Z}_2)\right) \to B_r\left(\ell^\infty(\mathbb{Z}_2)\right), \quad B_r\left(\ell^\infty(\mathbb{Z}_2)\right) := \{\mathbf{x} \in \ell^\infty(\mathbb{Z}_2) : \|\mathbf{x}\|_\infty \leqslant r\},$$

as long as $r \leqslant \tau$ and $|z| \leqslant c_\tau$ for some constant $c_\tau > 0$ depending only on c_1, c_2, δ_l and τ. Now it suffices to prove that h restricted to $B_r\left(\ell^\infty(\mathbb{Z}_2)\right)$ is a contraction, which implies that $\boldsymbol{\varepsilon} := \lim_{k\to\infty}\boldsymbol{\varepsilon}^{(k)}$ exists and is a unique solution to (C.14) subject to the condition $\|\boldsymbol{\varepsilon}\|_\infty \leqslant r$.

From the iteration relation (C.16), using (C.17) we obtain that

$$\boldsymbol{\varepsilon}^{(k+1)} - \boldsymbol{\varepsilon}^{(k)} = S^{-1}\left[\mathbf{e}(\boldsymbol{\varepsilon}^{(k)}) - \mathbf{e}(\boldsymbol{\varepsilon}^{(k-1)})\right] \leqslant \widetilde{C}\left\|\mathbf{e}(\boldsymbol{\varepsilon}^{(k)}) - \mathbf{e}(\boldsymbol{\varepsilon}^{(k-1)})\right\|_\infty. \tag{C.18}$$

From the expression of $\mathbf{e}$, we see that as long as r is chosen to be sufficiently small compared to $\theta_l^{-1} - g(m_{1c})g(m_{2c}) = (1-t_l)\theta_l^{-1}$, then

$$\left\|\mathbf{e}(\boldsymbol{\varepsilon}^{(k)}) - \mathbf{e}(\boldsymbol{\varepsilon}^{(k-1)})\right\|_\infty \leqslant C\left(\|\boldsymbol{\varepsilon}^{(k)}\|_\infty + \|\boldsymbol{\varepsilon}^{(k-1)}\|_\infty\right)\|\boldsymbol{\varepsilon}^{(k)} - \boldsymbol{\varepsilon}^{(k-1)}\|_\infty$$

for some constant $C > 0$ depending only on c_1, c_2 and δ_l. Thus we can choose a sufficiently small constant $0 < r \leqslant \min\{\tau, (2C)^{-1}\}$ such that $Cr \leqslant 1/2$. Then $\mathbf{f}$ is indeed a contraction mapping on $B_r(\ell^\infty(\mathbb{Z}_2))$, which proves both the existence and uniqueness of the solution to (C.14) if we choose c_0 in (C.10) as $c_0 = \min\{c_\tau, r\}$. After obtaining $\omega_{1c} = m_{1c} + \varepsilon_1$ and $\omega_{2c} = m_{2c} + \varepsilon_2$, we can define ω_{3c} and ω_{4c} using the first and third equations in (C.7).

Note that with (C.17) and $\boldsymbol{\varepsilon}^{(0)} = \mathbf{0}$, we get from (C.16) that $\|\boldsymbol{\varepsilon}^{(1)}\|_\infty \leqslant \widetilde{C}|z|$. Then with the contraction property of $\mathbf{f}$, we get that

$$\|\boldsymbol{\varepsilon}\|_\infty \leqslant \sum_{k=0}^{\infty} \|\boldsymbol{\varepsilon}^{(k+1)} - \boldsymbol{\varepsilon}^{(k)}\|_\infty \leqslant 2\widetilde{C}|z|.$$

This gives the bound (C.11) for ω_{1c} and ω_{2c}. Then using the first and third equations in (C.7), we immediately obtain the bound (C.11) for ω_{3c} and ω_{4c} as long as c_0 is sufficiently small. ☐

As a byproduct of the contraction mapping argument in the above proof, we also obtain the following stability result that will be used in the proof of Theorem C.4 below.

Lemma C.3. *There exist constants $c_0, C_0 > 0$ depending only on c_1, c_2 and δ_l such that the self-consistent equations in* (C.7) *are stable in the following sense. Suppose $|z| \leqslant c_0$ and $\omega_\alpha : \mathbb{C}_+ \mapsto \mathbb{C}_+$, $\alpha = 1, 2, 3, 4$, are analytic functions of z such that*

$$\max_{\alpha=1}^{4} |\omega_\alpha(z) - m_{\alpha c}(\theta_l)| \leqslant c_0. \tag{C.19}$$

Suppose they satisfy the system of equations

$$\begin{aligned} &\frac{c_1}{\omega_1} + z + \omega_3 = \mathcal{E}_1, \quad \omega_3 + (1-\theta_l)\frac{1 + (1-\theta_l)\omega_2}{[1 + (1-\theta_l)\omega_1][1 + (1-\theta_l)\omega_2] - \theta_l^{-1}} = \mathcal{E}_2, \\ &\frac{c_2}{\omega_2} + \omega_4 = \mathcal{E}_3, \quad \omega_4 + (1-\theta_l)\frac{1 + (1-\theta_l)\omega_1}{[1 + (1-\theta_l)\omega_1][1 + (1-\theta_l)\omega_2] - \theta_l^{-1}} = \mathcal{E}_4, \end{aligned} \tag{C.20}$$

for some errors bounded by $\max_{\alpha=1}^4 |\mathcal{E}_\alpha| \leqslant \delta(z)$, where $\delta(z)$ is a deterministic z-dependent function satisfying that $\delta(z) \leqslant (\log n)^{-1}$. Then we have

$$\max_{\alpha=1}^{4} |\omega_\alpha(z) - \omega_{\alpha c}(z)| \leqslant C_0\delta(z). \tag{C.21}$$

Proof. As in the proof of Lemma C.2, we subtract the equations (C.20) from (C.7), and consider the contraction principle for the functions $\varepsilon_\alpha(z) := \omega_\alpha(z) - \omega_{\alpha c}(z)$. We omit the details. ☐

The following theorem gives the anisotropic local law for $\mathcal{R}(z, 0)$.

Theorem C.4. *Suppose Assumption 2.1 holds. Then for any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$, the following anisotropic local law holds uniformly in $z \in \mathbf{D} := \{z \in \mathbb{C}_+ : |z| \leqslant (\log n)^{-1}\}$:*

$$|\langle \mathbf{u}, \mathcal{R}(z,0)\mathbf{v}\rangle - \langle \mathbf{u}, \Gamma(z)\mathbf{v}\rangle| \prec n^{-1/2}, \tag{C.22}$$

where $\Gamma(z)$ is defined in (C.9).

The proof of this theorem will be given in Section C.2 below. Now we use it to complete the proof of (A.24) when $\alpha = 1$.

Proof of (A.24) *for* GJ_1G. Using (C.3) and Cauchy's integral formula, we get that

$$\begin{aligned}\langle \mathbf{u}, G(\theta_l) J_\alpha G(\theta_l)\, \mathbf{v}\rangle &= \frac{1}{2\pi \mathrm{i}} \oint_{\mathcal{C}} \frac{\langle \mathbf{u}, \mathcal{R}(w,0)\, \mathbf{v}\rangle}{w^2} \mathrm{d}w = \frac{1}{2\pi \mathrm{i}} \oint_{\mathcal{C}} \frac{\langle \mathbf{u}, \Gamma(w)\, \mathbf{v}\rangle}{w^2} \mathrm{d}w + \mathrm{O}_\prec(n^{-1/2}) \\ &= \langle \mathbf{u}, \Gamma'(0)\, \mathbf{v}\rangle + \mathrm{O}_\prec(n^{-1/2}),\end{aligned} \tag{C.23}$$

where $\mathcal{C}$ is the contour $\{w \in \mathbb{C} : |w| = (\log n)^{-1}\}$ and we used (C.22) in the second step. It remains to calculate $\Gamma'(0)$, which is reduced to calculating the derivatives $\dot{m}_{\alpha c}(\theta_l) := \omega'_\alpha(z = 0)$, $\alpha = 1, 2, 3, 4$.

Using equation (C.7) and implicit differentiation, we obtain that

$$\begin{aligned}c_1^{-1}\dot{m}_{1c} &= m_{3c}^{-2} + \dot{m}_{1c} + \frac{\theta_l^{-1}}{[1+(1-\theta_l)m_{2c}]^2}\dot{m}_{2c}, \quad \dot{m}_{3c} = m_{3c}^2\left(c_1^{-1}\dot{m}_{1c} - m_{3c}^{-2}\right), \\ c_2^{-1}\dot{m}_{2c} &= \dot{m}_{2c} + \frac{\theta_l^{-1}}{[1+(1-\theta_l)m_{1c}]^2}\dot{m}_{1c}, \quad \dot{m}_{4c} = c_2^{-1}\dot{m}_{2c}m_{4c}^2.\end{aligned}$$

Solving the above equations and using that (recall equation (3.17))

$$\frac{\theta_l^{-1}}{[1+(1-\theta_l)m_{2c}]^2} = \frac{h^2}{m_{3c}^2}, \quad \frac{\theta_l^{-1}}{[1+(1-\theta_l)m_{1c}]^2} = \frac{h^2}{m_{4c}^2},$$

we get that $c_\alpha^{-1}\dot{m}_{\alpha c} = \gamma_\alpha^{(1)}$, $\alpha = 1, 2$, and $\dot{m}_{\alpha c} = \gamma_\alpha^{(1)}$, $\alpha = 3, 4$, for $\gamma_\alpha^{(1)}$ defined in (A.23). Moreover, we can check that $g_1'(0) = h_1(z)$. Hence we get $\Gamma'(0) = \Gamma^{(1)}(\theta_l)$, which, together with (C.23), concludes (A.24). □

The proof of Theorem A.1 for $GJ_\alpha G$ with $\alpha = 2, 3, 4$ is exactly the same, except that we need to use the following local law in Theorem C.5. Recall the resolvent $\mathcal{R}(w_1, w_2, w_3, w_4)$ defined in (C.2). We define $(\omega_{\alpha c}(\mathbf{w}))_{\alpha=1}^4$, as the unique solution to the following system of self-consistent equations

$$\begin{aligned}\frac{c_1}{\omega_{1c}} &= -w_1 - \omega_{3c}, \quad \frac{c_2}{\omega_{2c}} = -w_2 - \omega_{4c}, \\ \omega_{3c} &= (\theta_l - 1)\frac{1+(1-\theta_l)(\omega_{2c}+w_4)}{[1+(1-\theta_l)(\omega_{1c}+w_3)][1+(1-\theta_l)(\omega_{2c}+w_4)] - \theta_l^{-1}}, \\ \omega_{4c} &= (\theta_l - 1)\frac{1+(1-\theta_l)(\omega_{1c}+w_3)}{[1+(1-\theta_l)(\omega_{1c}+w_3)][1+(1-\theta_l)(\omega_{2c}+w_4)] - \theta_l^{-1}},\end{aligned} \tag{C.24}$$

such that $\operatorname{Im}\omega_{\alpha c}(\mathbf{w}) > 0$ whenever $\mathbf{w} \in \mathbb{C}_+^4$. Define the matrix limit of $\mathcal{R}(\mathbf{w})$ as

$$\Gamma(\mathbf{w}) := \begin{bmatrix} \begin{pmatrix} c_1^{-1}\omega_{1c}(\mathbf{w})I_p & 0 \\ 0 & c_2^{-1}\omega_{2c}(\mathbf{w})I_q \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} \omega_{3c}(\mathbf{w})I_n & \widetilde{g}(\mathbf{w})I_n \\ \widetilde{g}(\mathbf{w})I_n & \omega_{4c}(\mathbf{w})I_n \end{pmatrix} \end{bmatrix}, \tag{C.25}$$

where $\widetilde{g}(\mathbf{w})$ is defined by

$$\widetilde{g}(\mathbf{w}) := \frac{(\theta_l - 1)\theta_l^{-1/2}}{[1+(1-\theta_l)(\omega_{1c}+w_3)][1+(1-\theta_l)(\omega_{2c}+w_4)] - \theta_l^{-1}}. \tag{C.26}$$

Then we have the following local law on $\mathcal{R}(\mathbf{w})$.

Theorem C.5. *Suppose Assumption 2.1 holds. Fix any $\alpha = 1, 2, 3, 4$. For any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$, the following anisotropic local law holds uniformly in $w_\alpha \in \{w_\alpha \in \mathbb{C}_+ : |w_\alpha| \leqslant (\log n)^{-1}\}$ if $w_\beta = 0$ for $\beta \neq \alpha$:*

$$|\langle \mathbf{u}, \mathcal{R}(\mathbf{w})\mathbf{v}\rangle - \langle \mathbf{u}, \Gamma(\mathbf{w})\mathbf{v}\rangle| \prec n^{-1/2}. \tag{C.27}$$

This theorem can be proved in exactly the same way as Theorem C.4. Moreover, with Theorem C.5, the proof of Theorem A.1 for $GJ_\alpha G$, $\alpha = 2, 3, 4$, is also the same as the $\alpha = 1$ case. So we omit the details for both proofs.

C.2 Proof of Theorem C.4

In this section, we provide the proof of Theorem C.4. We first prove the following a priori estimates on $\mathcal{R}(z, 0)$. In the following proof, we will abbreviate $\mathcal{R}(z) \equiv \mathcal{R}(z, 0)$.

Lemma C.6. *There exists a constant $C > 0$ such that the following estimates hold with high probability:*

$$\sup_{z \in \mathbf{D}} \|\mathcal{R}(z)\| \leqslant C, \tag{C.28}$$

and

$$\sup_{z \in \mathbf{D}} \|\mathcal{R}(z) - G(\theta_l)\| \leqslant C|z|. \tag{C.29}$$

Proof. We denote the $(\mathcal{I}_1 \cup \mathcal{I}_2) \times (\mathcal{I}_1 \cup \mathcal{I}_2)$ block of $\mathcal{R}$ by $\mathcal{R}_L$, the $(\mathcal{I}_1 \cup \mathcal{I}_2) \times (\mathcal{I}_3 \cup \mathcal{I}_4)$ block by $\mathcal{R}_{LR}$, the $(\mathcal{I}_3 \cup \mathcal{I}_4) \times (\mathcal{I}_1 \cup \mathcal{I}_2)$ block by $\mathcal{R}_{RL}$, and the $(\mathcal{I}_3 \cup \mathcal{I}_4) \times (\mathcal{I}_3 \cup \mathcal{I}_4)$ block by $\mathcal{R}_R$. Using Schur complement formula, we obtain that

$$\mathcal{R}_L = \begin{pmatrix} \mathcal{R}_1 & -\theta_l^{-1/2}\mathcal{R}_1 S_{xy} S_{yy}^{-1} \\ -\theta_l^{-1/2} S_{yy}^{-1} S_{yx} \mathcal{R}_1 & \mathcal{R}_2 \end{pmatrix}, \tag{C.30}$$

where

$$\mathcal{R}_1 = \left(S_{xy} S_{yy}^{-1} S_{yx} - \theta_l S_{xx} - z\right)^{-1}, \quad \mathcal{R}_2 = -\theta_l^{-1} S_{yy}^{-1} + \theta_l^{-1} S_{yy}^{-1} S_{yx} \mathcal{R}_1 S_{xy} S_{yy}^{-1}.$$

The other blocks are given by

$$\mathcal{R}_R = \begin{pmatrix} \theta_l I_n & \theta_l^{1/2} I_n \\ \theta_l^{1/2} I_n & \theta_l I_n \end{pmatrix} + \begin{pmatrix} \theta_l I_n & \theta_l^{1/2} I_n \\ \theta_l^{1/2} I_n & \theta_l I_n \end{pmatrix} \begin{pmatrix} X^\top & 0 \\ 0 & Y^\top \end{pmatrix} \mathcal{R}_L \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \begin{pmatrix} \theta_l I_n & \theta_l^{1/2} I_n \\ \theta_l^{1/2} I_n & \theta_l I_n \end{pmatrix}, \tag{C.31}$$

and

$$\mathcal{R}_{LR} = -\mathcal{R}_L \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \begin{pmatrix} \theta_l I_n & \theta_l^{1/2} I_n \\ \theta_l^{1/2} I_n & \theta_l I_n \end{pmatrix}, \quad \mathcal{R}_{RL} = -\begin{pmatrix} \theta_l I_n & \theta_l^{1/2} I_n \\ \theta_l^{1/2} I_n & \theta_l I_n \end{pmatrix} \begin{pmatrix} X^\top & 0 \\ 0 & Y^\top \end{pmatrix} \mathcal{R}_L. \tag{C.32}$$

One can compare the above expressions with (6.44)–(6.46). With the estimates (6.10) and (6.11), we see that it suffices to prove the following estimates for $\mathcal{R}_1$:

$$\sup_{z \in \mathbf{D}} \|\mathcal{R}_1(z)\| \lesssim 1 \quad \text{with high probability}, \tag{C.33}$$

and

$$\sup_{z \in \mathbf{D}} \left\|\mathcal{R}_1(z) - \mathcal{G}_{(11)}(\theta_l)\right\| \lesssim |z| \quad \text{with high probability}, \tag{C.34}$$

where $\mathcal{G}_{(11)}$ is the $\mathcal{I}_1 \times \mathcal{I}_1$ block of G (recall the notations in Section 3.3). With $\mathcal{H}$ in (6.40), we can write $\mathcal{R}_1$ as

$$\mathcal{R}_1 = S_{xx}^{-1/2} \left(\mathcal{H}\mathcal{H}^\top - \theta_l - zS_{xx}^{-1}\right)^{-1} S_{xx}^{-1/2}.$$

By (3.4), we have that with high probability, $\theta_l - \mathcal{H}\mathcal{H}^\top$ is positive definite and its smallest eigenvalue satisfies

$$\lambda_p(\theta_l - \mathcal{H}\mathcal{H}^\top) \geqslant (\theta_l - \lambda_+)/2 \gtrsim 1.$$

Combining this estimate with (6.10), we obtain that with high probability,

$$\sup_{z \in \mathbf{D}} \|\mathcal{R}_1(z)\| \lesssim \frac{1}{\theta_l - \lambda_+ - \mathrm{O}((\log n)^{-1})} \lesssim 1.$$

This concludes (C.33). With (C.33), we can easily conclude (C.34):

$$\left|\langle \mathbf{u}, \mathcal{R}_1(z)\, \mathbf{v}\rangle - \langle \mathbf{u}, \mathcal{G}_{(11)}(\theta_l)\, \mathbf{v}\rangle\right| = |\langle \mathbf{u}, [\mathcal{R}_1(z) - \mathcal{R}_1(0)]\, \mathbf{v}\rangle| = |z|\, |\langle \mathbf{u}, \mathcal{R}_1(z)\mathcal{R}_1(0)\, \mathbf{v}\rangle| \lesssim |z|,$$

with high probability. ☐

Combining (C.29) with the local law (3.24), we can obtain the rough bound

$$\max_{z \in \mathbf{D}} \max_{\mathfrak{a},\mathfrak{b} \in \mathcal{I}} |\mathcal{R}_{\mathfrak{a}\mathfrak{b}}(z) - \Pi_{\mathfrak{a}\mathfrak{b}}(\theta_l)| \leqslant C(\log n)^{-1} \quad \text{with high probability.} \tag{C.35}$$

Then we record some useful resolvent identities in Lemma C.7 and Lemma C.8, which can be proved easily using Schur complement formula. For simplicity, we use the notations in (5.2).

Lemma C.7. *We have the following resolvent identities.*

(i) For $i \in \mathcal{I}_1 \cup \mathcal{I}_2$, we have that

$$\frac{1}{\mathcal{R}_{ii}} = -z\mathbf{1}_{i \in \mathcal{I}_1} - \left(W\mathcal{R}^{(i)}W^\top\right)_{ii}. \tag{C.36}$$

(ii) For $i \in \mathcal{I}_1 \cup \mathcal{I}_2$ and $\mathfrak{a} \in \mathcal{I}\backslash\{i\}$, we have that

$$\mathcal{R}_{i\mathfrak{a}} = -\mathcal{R}_{ii}\left(W\mathcal{R}^{(i)}\right)_{i\mathfrak{a}}. \tag{C.37}$$

(iii) For $\mathfrak{a} \in \mathcal{I}$ and $\mathfrak{b}, \mathfrak{c} \in \mathcal{I}\backslash\{\mathfrak{a}\}$, we have that

$$\mathcal{R}_{\mathfrak{b}\mathfrak{c}} = \mathcal{R}^{(\mathfrak{a})}_{\mathfrak{b}\mathfrak{c}} + \frac{\mathcal{R}_{\mathfrak{b}\mathfrak{a}}\mathcal{R}_{\mathfrak{a}\mathfrak{c}}}{\mathcal{R}_{\mathfrak{a}\mathfrak{a}}}. \tag{C.38}$$

(iv) All of the above identities hold for $\mathcal{R}^{(\mathbb{T})}$ instead of $\mathcal{R}$ for any index set $\mathbb{T} \subset \mathcal{I}$.

For $\mu, \nu \in \mathcal{I}_3$, we define the 2×2 blocks

$$\mathcal{R}_{[\mu\nu]} := \begin{pmatrix} \mathcal{R}_{\mu\nu} & \mathcal{R}_{\mu\overline{\nu}} \\ \mathcal{R}_{\overline{\mu}\nu} & \mathcal{R}_{\overline{\mu}\overline{\nu}} \end{pmatrix}, \tag{C.39}$$

where we denote $\overline{\mu} := \mu + n$ and $\overline{\nu} := \nu + n$. We call $\mathcal{R}_{[\mu\nu]}$ a diagonal block if $\mu = \nu$, and an off-diagonal block otherwise. For $i \in \mathcal{I}_1$, $j \in \mathcal{I}_2$ and $\mu \in \mathcal{I}_3$, we define the vectors

$$\mathcal{R}_{i,[\mu]} := (\mathcal{R}_{i\mu}, \mathcal{R}_{i\overline{\mu}}), \quad \mathcal{R}_{[\mu],i} := \begin{pmatrix} \mathcal{R}_{\mu i} \\ \mathcal{R}_{\overline{\mu} i} \end{pmatrix}. \tag{C.40}$$

For $\mu \in \mathcal{I}_3$, we denote $H^{[\mu]} := H^{(\mu\overline{\mu})}$ and $\mathcal{R}^{[\mu]} := \mathcal{R}^{(\mu\overline{\mu})}$ in the sense of Definition 4.2. Then we record the following resolvent identities, which can be obtained directly from Schur complement formula.

Lemma C.8. *We have the following resolvent identities.*

(i) For $\mu \in \mathcal{I}_3$, we have that

$$\mathcal{R}_{[\mu\mu]}^{-1} = \frac{1}{\theta_l - 1}\begin{pmatrix} 1 & -\theta_l^{-1/2} \\ -\theta_l^{-1/2} & 1 \end{pmatrix} - \begin{bmatrix} (X^\top \mathcal{R}^{[\mu]} X)_{\mu\mu} & (X^\top \mathcal{R}^{[\mu]} Y)_{\mu\overline{\mu}} \\ (Y^\top \mathcal{R}^{[\mu]} X)_{\overline{\mu}\mu} & (Y^\top \mathcal{R}^{[\mu]} Y)_{\overline{\mu}\overline{\mu}} \end{bmatrix}. \tag{C.41}$$

(ii) For $i \in \mathcal{I}_1 \cup \mathcal{I}_2$ and $\mu \in \mathcal{I}_3$, we have that

$$\mathcal{R}_{i,[\mu]} = \mathcal{R}_{[\mu],i}^\top = -\left[(\mathcal{R}^{[\mu]} X)_{i\mu}, (\mathcal{R}^{[\mu]} Y)_{i\overline{\mu}}\right] \mathcal{R}_{[\mu\mu]}. \tag{C.42}$$

(iii) For $\mu \neq \nu \in \mathcal{I}_3$, we have that

$$\mathcal{R}_{[\mu\nu]} = -\mathcal{R}_{[\mu\mu]} \begin{bmatrix} (X^\top \mathcal{R}^{[\mu]})_{\mu\nu} & (X^\top \mathcal{R}^{[\mu]})_{\mu\overline{\nu}} \\ (Y^\top \mathcal{R}^{[\mu]})_{\overline{\mu}\nu} & (Y^\top \mathcal{R}^{[\mu]})_{\overline{\mu}\overline{\nu}} \end{bmatrix} = -\begin{bmatrix} (\mathcal{R}^{[\nu]} X)_{\mu\nu} & (\mathcal{R}^{[\nu]} Y)_{\mu\overline{\nu}} \\ (\mathcal{R}^{[\nu]} X)_{\overline{\mu}\nu} & (\mathcal{R}^{[\nu]} Y)_{\overline{\mu}\overline{\nu}} \end{bmatrix} \mathcal{R}_{[\nu\nu]}. \tag{C.43}$$

(iv) For $\mu \in \mathcal{I}_3$ and $\mathfrak{a}_1, \mathfrak{a}_2, \mathfrak{b}_1, \mathfrak{b}_2 \in \mathcal{I} \backslash \{\mu, \overline{\mu}\}$, we have that

$$\begin{pmatrix} \mathcal{R}_{\mathfrak{a}_1\mathfrak{b}_1} & \mathcal{R}_{\mathfrak{a}_1\mathfrak{b}_2} \\ \mathcal{R}_{\mathfrak{a}_2\mathfrak{b}_1} & \mathcal{R}_{\mathfrak{a}_2\mathfrak{b}_2} \end{pmatrix} = \begin{pmatrix} \mathcal{R}^{[\mu]}_{\mathfrak{a}_1\mathfrak{b}_1} & \mathcal{R}^{[\mu]}_{\mathfrak{a}_1\mathfrak{b}_2} \\ \mathcal{R}^{[\mu]}_{\mathfrak{a}_2\mathfrak{b}_1} & \mathcal{R}^{[\mu]}_{\mathfrak{a}_2\mathfrak{b}_2} \end{pmatrix} + \begin{pmatrix} \mathcal{R}_{\mathfrak{a}_1\mu} & \mathcal{R}_{\mathfrak{a}_1\overline{\mu}} \\ \mathcal{R}_{\mathfrak{a}_2\mu} & \mathcal{R}_{\mathfrak{a}_2\overline{\mu}} \end{pmatrix} \mathcal{R}_{[\mu\mu]}^{-1} \begin{pmatrix} \mathcal{R}_{\mu\mathfrak{b}_1} & \mathcal{R}_{\mu\mathfrak{b}_2} \\ \mathcal{R}_{\overline{\mu}\mathfrak{b}_1} & \mathcal{R}_{\overline{\mu}\mathfrak{b}_2} \end{pmatrix}. \tag{C.44}$$

Using the above tools, we now prove the following entrywise version of Theorem C.4.

Proposition C.9 (Entrywise local law)**.** *If Assumption 2.1 holds, then we have that*

$$\max_{\mathfrak{a},\mathfrak{b} \in \mathcal{I}} |\mathcal{R}_{\mathfrak{a}\mathfrak{b}}(z, 0) - \Gamma_{\mathfrak{a}\mathfrak{b}}(z)| \prec n^{-1/2} \quad \textit{uniformly in } z \in \mathbf{D}. \tag{C.45}$$

For the proof of Proposition C.9, we introduce the following $\mathcal{Z}$ variables

$$\mathcal{Z}_{\mathfrak{a}} := (1 - \mathbb{E}_{\mathfrak{a}})\big(\mathcal{R}_{\mathfrak{a}\mathfrak{a}}\big)^{-1},$$

where $\mathbb{E}_{\mathfrak{a}}[\cdot] := \mathbb{E}[\cdot \mid H^{(\mathfrak{a})}]$, i.e. it is the partial expectation over the randomness of the $\mathfrak{a}$-th row and column of H. By (C.36), we have that for $i \in \mathcal{I}_\alpha$, $\alpha = 1, 2$,

$$\mathcal{Z}_i = (\mathbb{E}_i - 1)\left(W \mathcal{R}^{(i)} W^\top\right)_{ii} = \sum_{\mu,\nu \in \mathcal{I}_{\alpha+2}}^{(i)} \mathcal{R}^{(i)}_{\mu\nu} \left(\frac{1}{n}\delta_{\mu\nu} - W_{i\mu} W_{i\nu}\right). \tag{C.46}$$

We also introduce the matrix-valued $\mathcal{Z}$ variables

$$\mathcal{Z}_{[\mu]} := \left(1 - \mathbb{E}_{[\mu]}\right)\left(\mathcal{R}_{[\mu\mu]}\right)^{-1}, \tag{C.47}$$

where $\mathbb{E}_{[\mu]}[\cdot] := \mathbb{E}[\cdot \mid H^{[\mu]}]$, i.e. it is the partial expectation over the randomness of the μ-th and $\overline{\mu}$-th rows and columns of H. By (C.41), we have that

$$\mathcal{Z}_{[\mu]} = \begin{bmatrix} \sum_{i,j \in \mathcal{I}_1} \mathcal{R}^{[\mu]}_{ij} (n^{-1}\delta_{ij} - X_{i\mu} X_{j\mu}) & \sum_{i \in \mathcal{I}_1, j \in \mathcal{I}_2} \mathcal{R}^{[\mu]}_{ij} X_{i\mu} Y_{j\overline{\mu}} \\ \sum_{i \in \mathcal{I}_1, j \in \mathcal{I}_2} \mathcal{R}^{[\mu]}_{ji} X_{i\mu} Y_{j\overline{\mu}} & \sum_{i,j \in \mathcal{I}_2} \mathcal{R}^{[\mu]}_{ij} (n^{-1}\delta_{ij} - Y_{i\overline{\mu}} Y_{j\overline{\mu}}) \end{bmatrix}. \tag{C.48}$$

We also define the random error to control the off-diagonal entries,

$$\Lambda_o := \max_{i \neq j \in \mathcal{I}_1 \cup \mathcal{I}_2} |\mathcal{R}_{ij}| + \max_{\mu \neq \nu \in \mathcal{I}_3} \|\mathcal{R}_{[\mu\nu]}\| + \max_{i \in \mathcal{I}_1 \cup \mathcal{I}_2, \mu \in \mathcal{I}_3} \left\|\mathcal{R}_{i,[\mu]}\right\|. \tag{C.49}$$

Now we claim the following large deviation estimate.

Claim C.10. *Under the setting of Theorem C.4, we have that*

$$\Lambda_o + |Z_i| + \|Z_{[\mu]}\| \prec n^{-1/2}. \tag{C.50}$$

Proof. For $i \in \mathcal{I}_\alpha$, $\alpha = 1, 2$, applying Lemma 4.3 to $\mathcal{Z}_i$ in (C.46), we get that

$$|\mathcal{Z}_i| \prec \frac{1}{n}\Big(\sum_{\mu,\nu \in \mathcal{I}_{\alpha+2}} |\mathcal{R}^{(i)}_{\mu\nu}|^2 \Big)^{1/2} \leqslant \frac{1}{\sqrt{n}} \Big[\frac{1}{n} \sum_{\mu \in \mathcal{I}_{\alpha+2}} \Big(\mathcal{R}^{(i)} (\mathcal{R}^{(i)})^* \Big)_{\mu\mu} \Big]^{1/2} \prec n^{-1/2},$$

where in the last step we applied (C.28) to $\mathcal{R}^{(i)}$ to get $(\mathcal{R}^{(i)}(\mathcal{R}^{(i)})^*)_{\mu\mu} = \mathrm{O}(1)$ with high probability (because $\mathcal{R}^{(i)}$ satisfies the same assumption as $\mathcal{R}$). Similarly, applying Lemma 4.3 to $\mathcal{Z}_{[\mu]}$ in (C.48), we obtain that

$$\|\mathcal{Z}_{[\mu]}\| \prec \frac{1}{n}\Big(\sum_{i,j \in \mathcal{I}_1 \cup \mathcal{I}_2} |\mathcal{R}^{[\mu]}_{ij}|^2 \Big)^{1/2} = \frac{1}{\sqrt{n}} \Big[\frac{1}{n} \sum_{i \in \mathcal{I}_1 \cup \mathcal{I}_2} \Big(\mathcal{R}^{[\mu]} (\mathcal{R}^{[\mu]})^* \Big)_{ii} \Big]^{1/2} \prec n^{-1/2}. \tag{C.51}$$

The proof of the off-diagonal estimate is similar. For $i \neq j \in \mathcal{I}_1 \cup \mathcal{I}_2$, using (C.37), Lemma 4.3 and (C.28), we obtain that

$$|\mathcal{R}_{ij}| \prec \frac{1}{\sqrt{n}} \Big(\sum_{\mu \in \mathcal{I}_3 \cup \mathcal{I}_4} |\mathcal{R}^{(i)}_{\mu j}|^2 \Big)^{1/2} \prec n^{-1/2}.$$

For $\mu \neq \nu \in \mathcal{I}_3$, using (C.43), Lemma 4.3 and (C.28) we obtain that

$$\|\mathcal{R}_{[\mu\nu]}\| \prec \frac{1}{n}\Big(\sum_{i \in \mathcal{I}_1 \cup \mathcal{I}_2} |\mathcal{R}^{[\mu]}_{i\nu}|^2 \Big)^{1/2} + \frac{1}{n}\Big(\sum_{i \in \mathcal{I}_1 \cup \mathcal{I}_2} |\mathcal{R}^{[\mu]}_{i\bar{\nu}}|^2 \Big)^{1/2} \prec n^{-1/2}.$$

Finally, for $i \in \mathcal{I}_1 \cup \mathcal{I}_2$ and $\mu \in \mathcal{I}_3$, using (C.42), Lemma 4.3 and (C.28), we obtain that

$$\|\mathcal{R}_{i,[\mu]}\| \prec \frac{1}{n}\Big(\sum_{j \in \mathcal{I}_1 \cup \mathcal{I}_2} |\mathcal{R}^{[\mu]}_{ij}|^2 \Big)^{1/2} \prec n^{-1/2}.$$

Combining the above estimates, we conclude (C.50). □

A key component of the proof for Proposition C.9 is to show that ω_α, $\alpha = 1, 2, 3, 4$, satisfy the self-consistent equations in (C.20) up to some small errors $|\mathcal{E}_\alpha| \prec n^{-1/2}$.

Lemma C.11. *Fix any constant $\varepsilon > 0$. The following estimates hold uniformly in $z \in \mathbf{D}$:*

$$\left| \frac{c_1}{\omega_1} + z + \omega_3 \right| \prec n^{-1/2}, \quad \left| \frac{c_2}{\omega_2} + \omega_4 \right| \prec n^{-1/2}, \tag{C.52}$$

$$\left| \omega_3 + (1-\theta_l) \frac{1 + (1-\theta_l)\omega_2}{[1 + (1-\theta_l)\omega_1][1 + (1-\theta_l)\omega_2] - \theta_l^{-1}} \right| \prec n^{-1/2}, \tag{C.53}$$

$$\left| \omega_4 + (1-\theta_l) \frac{1 + (1-\theta_l)\omega_1}{[1 + (1-\theta_l)\omega_1][1 + (1-\theta_l)\omega_2] - \theta_l^{-1}} \right| \prec n^{-1/2}. \tag{C.54}$$

Proof. Similar to (C.5), for $i \in \mathcal{I}_1 \cup \mathcal{I}_2$ and $\mu \in \mathcal{I}_3$, we denote

$$\omega_\alpha^{(i)} := \frac{1}{n} \sum_{\mathfrak{a} \in \mathcal{I}_\alpha} \mathcal{R}^{(i)}_{\mathfrak{a}\mathfrak{a}}, \quad \omega_\alpha^{[\mu]} := \frac{1}{n} \sum_{i \in \mathcal{I}_\alpha} \mathcal{R}^{[\mu]}_{\mathfrak{a}\mathfrak{a}}, \quad \alpha = 1, 2, 3, 4.$$

Using (C.36) and (C.46), we get that for $i \in \mathcal{I}_1$ and $j \in \mathcal{I}_2$,

$$\frac{1}{\mathcal{R}_{ii}} = -z - \omega_3 + \varepsilon_i, \quad \frac{1}{\mathcal{R}_{jj}} = -\omega_4 + \varepsilon_j, \tag{C.55}$$

where

$$\varepsilon_i := \mathcal{Z}_i + \omega_3 - \omega_3^{(i)}, \quad \varepsilon_j := \mathcal{Z}_j + \omega_4 - \omega_4^{(j)}.$$

On the other hand, using (C.41) and (C.47), we get that for $\mu \in \mathcal{I}_3$,

$$\mathcal{R}_{[\mu\mu]}^{-1} = \frac{1}{\theta_l - 1}\begin{pmatrix} 1 & -\theta_l^{-1/2} \\ -\theta_l^{-1/2} & 1 \end{pmatrix} - \begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix} + \varepsilon_\mu, \tag{C.56}$$

where

$$\varepsilon_\mu := \mathcal{Z}_\mu + \begin{pmatrix} \omega_1 - \omega_1^{[\mu]} & 0 \\ 0 & \omega_2 - \omega_2^{[\mu]} \end{pmatrix}.$$

Now using (C.38) and (C.50), we get that

$$\omega_3 - \omega_3^{(i)} = \frac{1}{n}\sum_{\mu\in\mathcal{I}_3} \frac{\mathcal{R}_{\mu i}\mathcal{R}_{i\mu}}{\mathcal{R}_{ii}} = \mathrm{O}_\prec(n^{-1}),$$

where in the second step we also used $|\mathcal{R}_{ii}| \gtrsim 1$ by (C.35) and (C.1). We have similar estimates for $\omega_4 - \omega_4^{(j)}$, $\omega_1 - \omega_1^{[\mu]}$ and $\omega_2 - \omega_2^{[\mu]}$. Together with (C.50), these estimates give that

$$\max_{i\in\mathcal{I}_1\cup\mathcal{I}_2} |\varepsilon_i| + \max_{\mu\in\mathcal{I}_3} \|\varepsilon_\mu\| \prec n^{-1/2}. \tag{C.57}$$

Now using the first equation in (C.55) and (C.57), we obtain that

$$\omega_1 = \frac{1}{n}\sum_{i\in\mathcal{I}_1} \mathcal{R}_{ii} = \frac{1}{n}\sum_{i\in\mathcal{I}_1} \frac{1}{-z-\omega_3+\varepsilon_i} = \frac{c_1}{-z-\omega_3} + \mathrm{O}_\prec(n^{-1/2}), \tag{C.58}$$

where in the second step we used $|z + \omega_3| \gtrsim 1$ with high probability by (C.35). This gives the first equation in (C.52). Similarly, using the second equation in (C.55), we can obtain the second equation in (C.52). With (C.35) and (C.1), we can check that

$$\left\| \left[\frac{1}{\theta_l - 1}\begin{pmatrix} 1 & -\theta_l^{-1/2} \\ -\theta_l^{-1/2} & 1 \end{pmatrix} - \begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix} \right]^{-1} \right\| \lesssim 1 \quad \text{with high probability.} \tag{C.59}$$

Taking the matrix inverse of (C.56) and using (C.57) and (C.59), we obtain that for $\mu \in \mathcal{I}_3$,

$$\mathcal{R}_{[\mu\mu]} = \frac{\theta_l - 1}{[1+(1-\theta_l)\omega_1][1+(1-\theta_l)\omega_2] - \theta_l^{-1}} \begin{pmatrix} 1+(1-\theta_l)\omega_2 & \theta_l^{-1/2} \\ \theta_l^{-1/2} & 1+(1-\theta_l)\omega_1 \end{pmatrix} + \mathrm{O}_\prec(n^{-1/2}). \tag{C.60}$$

After taking the average $n^{-1}\sum_{\mu\in\mathcal{I}_3}$ over the $(1,1)$-th and $(2,2)$-th entries of equation (C.60), we obtain the equations (C.53) and (C.54). □

Proo of Proposition C.9. We can apply Lemma C.3 now, where (C.19) is implied by (C.35), and the equations in (C.20) follow from Lemma C.11. Then (C.21) implies that

$$\max_{\alpha=1}^{4} |\omega_\alpha(z) - \omega_{\alpha c}(z)| \prec n^{-1/2}. \tag{C.61}$$

Plugging (C.61) into (C.55) and (C.60), we then get the diagonal estimate

$$\max_{i \in \mathcal{I}_1} \left|\mathcal{R}_{ii} - c_1^{-1}\omega_{1c}\right| + \max_{j \in \mathcal{I}_2} \left|\mathcal{R}_{jj} - c_2^{-1}\omega_{2c}\right| + \max_{\mu \in \mathcal{I}_3} \left\| \mathcal{R}_{[\mu\mu]} - \begin{pmatrix} \omega_{3c} & g_1 \\ g_1 & \omega_{4c} \end{pmatrix} \right\| \prec n^{-1/2}.$$

Combining it with the off-diagonal estimate in (C.50), we conclude (C.45). ☐

Finally, we can complete the proof of Theorem C.4.

Proof of Theorem C.4. With the entrywise local law, Proposition C.9, the proof of (C.22) uses a polynomialization method developed in [8]. Actually the argument is exactly the same as the one in Section 7 of [40]. Hence we omit the details. However, we make one remark that in the proof, we need to bound the high moments

$$\mathbb{E}\left|\langle \mathbf{u}, \mathcal{R}(z,0)\mathbf{v}\rangle - \langle \mathbf{u}, \Gamma(z)\mathbf{v}\rangle\right|^{2a}$$

for any fixed large $a \in \mathbb{N}$. Hence for regularity reasons, we shall use the resolvent $\mathcal{R}(z + \mathrm{i}n^{-4}, z')$ with $z' = \mathrm{i}n^{-4}$ in order to make use of the deterministic bound (C.6) on exceptional low-probability events, which justifies the applicability of Lemma 3.2 (iii). The structure of the proof is as follows. First, the argument in the proof of Claim 5.2 allows us to extend the entrywise local law (C.45) to $\mathcal{R}(z + \mathrm{i}n^{-4}, z')$. Then we can prove the anisotropic local law (C.22) for $\mathcal{R}(z + \mathrm{i}n^{-4}, z')$ using the arguments in Section 7 of [40]. After that, applying the argument in the proof of Claim 5.2 again allows us to extend the anisotropic local law to $\mathcal{R}(z, 0)$. ☐

References

[1] Zhidong Bai and Jianfeng Yao. Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. H. Poincaré Probab. Statist.*, 44(3):447–474, 2008.

[2] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

[3] Jinho Baik and Jack W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382 – 1408, 2006.

[4] Zhigang Bao, Jiang Hu, Guangming Pan, and Wang Zhou. Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case. *Ann. Statist.*, 47(1):612–640, 2019.

[5] Serban T. Belinschi, Hari Bercovici, Mireille Capitaine, and Maxime Février. Outliers in the spectrum of large deformed unitarily invariant models. *Ann. Probab.*, 45(6A):3571–3625, 2017.

[6] Florent Benaych-Georges, Alice Guionnet, and Mylène Maida. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.*, 16:1621–1662, 2011.

[7] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494 – 521, 2011.

[8] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19(33):1–53, 2014.

[9] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin. On the principal components of sample covariance matrices. *Prob. Theor. Rel. Fields*, 164(1):459–552, 2016.

[10] M. Capitaine, C. Donati-Martin, and D. Féral. Central limit theorems for eigenvalues of deformations of Wigner matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 48(1):107–133, 2012.

[11] Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.*, 37(1):1–47, 2009.

[12] Xiucai Ding and Fan Yang. Spiked separable covariance matrices and principal components. *The Annals of Statistics*, 49(2):1113 – 1138, 2021.

[13] L. Erdős, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré*, 14:1837–1926, 2013.

[14] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Delocalization and diffusion profile for random band matrices. *Commun. Math. Phys.*, 323:367–416, 2013.

[15] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős-Rényi graphs I: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 2013.

[16] Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Communications in Mathematical Physics*, 272(1):185–228, 2007.

[17] Delphine Féral and Sandrine Péché. The largest eigenvalues of sample covariance matrices for a spiked population: Diagonal case. *Journal of Mathematical Physics*, 50(7):073302, 2009.

[18] P.J. Forrester. The spectrum edge of random matrix ensembles. *Nucl. Phys. B*, 402(3):709 – 728, 1993.

[19] Yasunori Fujikoshi. High-dimensional asymptotic distributions of characteristic roots in multivariate linear models and canonical correlation analysis. *Hiroshima Math. J.*, 47(3):249–271, 2017.

[20] Xiao Han, Guangming Pan, and Qing Yang. A unified matrix model including both CCA and F matrices in multivariate analysis: The largest eigenvalue and its applications. *Bernoulli*, 24(4B):3447–3468, 2018.

[21] Xiao Han, Guangming Pan, and Bo Zhang. The Tracy-Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.*, 44(4):1564–1592, 2016.

[22] Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3-4):321–377, 1936.

[23] Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29:295–327, 2001.

[24] Iain M. Johnstone. Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Statist.*, 36(6):2638–2716, 2008.

[25] Alexei M. Khorunzhy, Boris A. Khoruzhenko, and Leonid A. Pastur. Asymptotic properties of large random matrices with independent entries. *Journal of Mathematical Physics*, 37(10):5033–5060, 1996.

[26] Antti Knowles and Jun Yin. The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.*, 66:1663–1749, 2013.

[27] Antti Knowles and Jun Yin. The outliers of a deformed Wigner matrix. *Ann. Probab.*, 42(5):1980–2031, 2014.

[28] Antti Knowles and Jun Yin. Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, pages 1–96, 2016.

[29] Ji Oon Lee and Jun Yin. A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.*, 163:117–173, 2014.

[30] A. Lytova and L. Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, 37(5):1778–1840, 09 2009.

[31] Zongming Ma and Fan Yang. Sample canonical correlation coefficients of high-dimensional random vectors with finite rank correlations. *arXiv:2102.03297*, 2021.

[32] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, 2007.

[33] S. Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probability Theory and Related Fields*, 134(1):174–174, 2006.

[34] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, pages 583–602. University of California Press, 1972.

[35] Craig A. Tracy and Harold Widom. Level-spacing distributions and the airy kernel. *Comm. Math. Phys.*, 159:151–174, 1994.

[36] Craig A. Tracy and Harold Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177:727–754, 1996.

[37] Kenneth W. Wachter. The limiting empirical measure of multiple discriminant ratios. *Ann. Statist.*, 8(5):937–957, 1980.

[38] Qinwen Wang and Jianfeng Yao. Extreme eigenvalues of large-dimensional spiked Fisher matrices with application. *Ann. Statist.*, 45(1):415–460, 2017.

[39] Fan Yang. Edge universality of separable covariance matrices. *Electron. J. Probab.*, 24:57 pp., 2019.

[40] Fan Yang. Sample canonical correlation coefficients of high-dimensional random vectors: local law and Tracy-Widom limit. *arXiv:2002.09643*, 2020.

[41] Fan Yang, Sifan Liu, Edgar Dobriban, and David P. Woodruff. How to reduce dimension with PCA and random projections? *arXiv:2005.00511.*

[42] Yanrong Yang and Guangming Pan. The convergence of the empirical distribution of canonical correlation coefficients. *Electron. J. Probab.*, 17:13 pp., 2012.

[43] Yanrong Yang and Guangming Pan. Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann. Statist.*, 43(2):467–500, 04 2015.